Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 5 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Regulatory Frameworks for Clinical Trial Data Sharing: Scoping Review (e33591)
Nachiket Gudi, Prashanthi Kamath, Trishnika Chakraborty, Anil Jacob, Shradha Parsekar, Suptendra Sarbadhikari, Oommen John. 7

Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review (e30405)
Timothy Kariotis, Megan Prictor, Shanton Chang, Kathleen Gray. 23

eHealth Tools That Assess and Track Health and Well-being in Children and Young People: Systematic Review (e26015)
Elizabeth Stewart, Alyssa Milton, Hannah Yee, Michael Song, Anna Roberts, Tracey Davenport, Ian Hickie. 47

Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review (e35371)
Robert Jakob, Samira Harperink, Aaron Rudolf, Elgar Fleisch, Severin Haug, Jacqueline Mair, Alicia Salamanca-Sanabria, Tobias Kowatsch.
9 3

Findings and Guidelines on Provider Technology, Fatigue, and Well-being: Scoping Review (e34451)
Donald Hilty, Christina Armstrong, Shelby Smout, Allison Crawford, Marlene Maheu, Kenneth Drude, Steven Chan, Peter Yellowlees, Elizabeth Krupinski. 119

Comparison of the ACASI Mode to Other Survey Modes in Sexual Behavior Surveys in Asia and Sub-Saharan Africa: Systematic Literature Review (e37356)
Nang Phoo, Roanna Lobo, Daniel Vujcich, Alison Reid. 134

Original Papers

Use of Robots in Critical Care: Systematic Review (e33380)
Rachel Teng, Yichen Ding, Kay See. 69

Delivering Mindfulness-Based Interventions for Insomnia, Pain, and Dysfunctional Eating Through a Text Messaging App: Three Randomized Controlled Trials Investigating the Effectiveness and Mediating Mechanisms (e30073)
Amanda Li, Keith Wong, Floria Chio, Winnie Mak, Loretta Poon. 187

The Effects of a Web-Based Tool for Parents of Children With Juvenile Idiopathic Arthritis: Randomized Controlled Trial (e29787)
Kathleen Mulligan, Shashivadan Hirani, Sally Harris, Jo Taylor, Lucy Wedderburn, Stanton Newman, WebParC Investigator group. 218

XSL•FO
RenderX

A Gamified Personalized Normative Feedback App to Reduce Drinking Among Sexual Minority Women: Randomized Controlled Trial and Feasibility Study (e34853)
Sarah Boyle, Joseph LaBrie, Bradley Trager, Lauren Costine. . . . 239

Telephone and Web-Based Delivery of Healthy Eating and Active Living Interventions for Parents of Children Aged 2 to 6 Years: Mixed Methods Process Evaluation of the Time for Healthy Habits Translation Trial (e35771)
Megan Hammersley, Rebecca Wyse, Rachel Jones, Anthony Okely, Luke Wolfenden, Simon Eckermann, Joe Xu, Amanda Green, Fiona Stacey, Sze Yoong, Jacklyn Jackson, Christine Innes-Hughes, Vincy Li, Chris Rissel. . . . 259

Disparity and Factors Associated With Internet Health Information Seeking Among US Adults Living With Diabetes Mellitus: Cross-sectional Study (e32723)
Ransome Eke, Khadiza Jannat, Xin Yang, Jason Parton. . . . 273

Comparison Between the 24-hour Holter Test and 72-hour Single-Lead Electrocardiogram Monitoring With an Adhesive Patch-Type Device for Atrial Fibrillation Detection: Prospective Cohort Study (e37970)
Soonil Kwon, So-Ryoung Lee, Eue-Keun Choi, Hyo-Jeong Ahn, Hee-Seok Song, Young-Shin Lee, Seil Oh, Gregory Lip. . . . 288

The Influence of Perceived Social Presence on the Willingness to Communicate in Mobile Medical Consultations: Experimental Study (e31797)
Lijuan Chen, Danyang Zhang, Mutian Hou. . . . 299

Screening Depressive Symptoms and Incident Major Depressive Disorder Among Chinese Community Residents Using a Mobile App–Based Integrated Mental Health Care Model: Cohort Study (e30907)
Huimin Zhang, Yuhua Liao, Xue Han, Beifang Fan, Yifeng Liu, Leanna Lui, Yena Lee, Mehala Subramaniapillai, Lingjiang Li, Lan Guo, Ciyong Lu, Roger McIntyre. . . . 316

The Use of Social Media as a Persuasive Platform to Facilitate Nutrition and Health Behavior Change in Young Adults: Web-Based Conversation Study (e28063)
Vanessa Friedman, Cassandra Wright, Annika Molenaar, Tracy McCaffrey, Linda Brennan, Megan Lim. . . . 331

Characterization of False or Misleading Fluoride Content on Instagram: Infodemiology Study (e37519)
Matheus Lotto, Tamires Sá Menezes, Irfhana Zakir Hussain, Shu-Feng Tsao, Zahid Ahmad Butt, Plinio P Morita, Thiago Cruvinel. . . . 346

Patient Activeness During Online Medical Consultation in China: Multilevel Analysis (e35557)
Bolin Cao, Wensen Huang, Naipeng Chao, Guang Yang, Ningzheng Luo. . . . 359

A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study (e33507)
Géraldine Falgarone, Guilhem Bousquet, Arnaud Wilmet, Albert Brizio, Valérie Faure, Celestin Guillouet, Franck Baudino, Isabelle Roque, Samuel Mayol, Frederic Pamoukdjian. . . . 370

Online Video Teletherapy Treatment of Obsessive-Compulsive Disorder Using Exposure and Response Prevention: Clinical Outcomes From a Retrospective Longitudinal Observational Study (e36431)
Jamie Feusner, Nicholas Farrell, Jeremy Kreyling, Patrick McGrath, Andreas Rhode, Ted Faneuff, Stephanie Lonsway, Reza Mohideen, John Jurich, Larry Trusky, Stephen Smith. . . . 381

How We Can Reap the Full Benefit of Teleconsultations: Economic Evaluation Combined With a Performance Evaluation Through a Discrete-Event Simulation (e32002)
Marius Huguet, Marianne Sarazin, Lionel Perrier, Vincent Augusto. . . . 393

Health Care Professionals' Experiences of Web-Based Symptom Checkers for Triage: Cross-sectional Survey Study (e33505)
Sari Kujala, Iiris Hörhammer. . . . 415

XSL•FO
RenderX

Promised and Lottery Airtime Incentives to Improve Interactive Voice Response Survey Participation Among Adults in Bangladesh and Uganda: Randomized Controlled Trial (e36943)
Dustin Gibson, Gulam Kibria, George Pariyo, Saifuddin Ahmed, Joseph Ali, Alain Labrique, Iqbal Khan, Elizeus Rutebemberwa, Meerjady Flora, Adnan Hyder. . . . 426

Self-administered Web-Based Tests of Executive Functioning and Perceptual Speed: Measurement Development Study With a Large Probability-Based Survey Panel (e34347)
Ying Liu, Stefan Schneider, Bart Orriens, Erik Meijer, Jill Darling, Tania Gutsche, Margaret Gatz. . . . 438

Access to and Use of Internet and Social Media by Low-Morbidity Stroke Survivors Participating in a National Web-Based Secondary Stroke Prevention Trial: Cross-sectional Survey (e33291)
Brigid Clancy, Billie Bonevski, Coralie English, Amanda Baker, Alyna Turner, Parker Magin, Michael Pollack, Robin Callister, Ashleigh Guillaumier. . . . 454

Self-efficacy and Use of Digital Health Care and Social Welfare Services Among Incarcerated People: Cross-sectional Survey Study (e36799)
Teemu Rantanen, Eeva Järveläinen, Teppo Leppälahti. . . . 467

Web-Based Independent Versus Laboratory-Based Stop-Signal Task Performance: Within-Subjects Counterbalanced Comparison Study (e32922)
Antoinette Poulton, Li Chen, Gezelle Dali, Michael Fox, Robert Hester. . . . 480

The Development and Evaluation of an e-Learning Course That Promotes Digital Health Literacy in School-age Children: Pre-Post Measurement Study (e37523)
Lars König, Eugenia Marbach-Breitrück, Anne Engler, Ralf Suhr. . . . 492

A Short Intervention and an Interactive e-Learning Module to Motivate Medical and Dental Students to Enlist as First Responders: Implementation Study (e38508)
Victor Taramarcaz, Tara Herren, Eric Golay, Simon Regard, Sébastien Martin-Achard, Francois Mach, Nicolas Schnetzler, Gaëtan Ricci, Ido Zamberg, Robert Larribau, Marc Niquille, Mélanie Suppan, Eduardo Schiffer, Laurent Suppan. . . . 508

Exploring Online Peer Support Groups for Adults Experiencing Long COVID in the United Kingdom: Qualitative Interview Study (e37674)
Hannah Day. . . . 522

Integration of an Intensive Care Unit Visualization Dashboard (i-Dashboard) as a Platform to Facilitate Multidisciplinary Rounds: Cluster-Randomized Controlled Trial (e35981)
Chao-Han Lai, Kai-Wen Li, Fang-Wen Hu, Pei-Fang Su, I-Lin Hsu, Min-Hsin Huang, Yen-Ta Huang, Ping-Yen Liu, Meng-Ru Shen. . . . 537

Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework (e31231)
Bettina Zimmermann, Theresa Willem, Carl Bredthauer, Alena Buyx. . . . 551

Distinguishing Admissions Specifically for COVID-19 From Incidental SARS-CoV-2 Admissions: National Retrospective Electronic Health Record Study (e37931)
Jeffrey Klann, Zachary Strasser, Meghan Hutch, Chris Kennedy, Jayson Marwaha, Michele Morris, Malarkodi Samayamuthu, Ashley Pfaff, Hossein Estiri, Andrew South, Griffin Weber, William Yuan, Paul Avillach, Kavishwar Wagholikar, Yuan Luo, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Gilbert Omenn, Shyam Visweswaran, John Holmes, Zongqi Xia, Gabriel Brat, Shawn Murphy. . . . 568

The Effect of Dissemination Pathways on Uptake and Relative Costs for a Transdiagnostic, Self-guided Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation: Comparative Implementation Study (e34769)
Philip Batterham, Amelia Gulliver, Ella Kurz, Louise Farrer, Christiaan Vis, Josien Schuurmans, Alison Calear. . . . 587

XSL•FO
RenderX

Factors Predicting Engagement of Older Adults With a Coach-Supported eHealth Intervention Promoting Lifestyle Change and Associations Between Engagement and Changes in Cardiovascular and Dementia Risk: Secondary Analysis of an 18-Month Multinational Randomized Controlled Trial (e32006)
Nicola Coley, Laurine Andre, Marieke Hoevenaar-Blom, Tiia Ngandu, Cathrien Beishuizen, Mariagnese Barbera, Lennard van Wanrooij, Miia Kivipelto, Hilkka Soininen, Willem van Gool, Carol Brayne, Eric Moll van Charante, Edo Richard, Sandrine Andrieu, HATICE study group, PRODEMOS study group. . . . 599

Readability of English, German, and Russian Disease-Related Wikipedia Pages: Automated Computational Analysis (e36835)
Jelizaveta Gordejeva, Richard Zowalla, Monika Pobiruchin, Martin Wiesner. . . . 615

The Effectiveness of a Computer-Tailored Web-Based Physical Activity Intervention Using Fitbit Activity Trackers in Older Adults (Active for Life): Randomized Controlled Trial (e31352)
Stephanie Alley, Jannique van Uffelen, Stephanie Schoeppe, Lynne Parkinson, Susan Hunt, Deborah Power, Natasha Waterman, Courtney Waterman, Quyen To, Mitch Duncan, Anthony Schneiders, Corneel Vandelanotte. . . . 638

The Accuracy of Artificial Intelligence in the Endoscopic Diagnosis of Early Gastric Cancer: Pooled Analysis Study (e27694)
Pei-Chin Chen, Yun-Ru Lu, Yi-No Kang, Chun-Chao Chang. . . . 655

Impact of Hospital Characteristics and Governance Structure on the Adoption of Tracking Technologies for Clinical and Supply Chain Use: Longitudinal Study of US Hospitals (e33742)
Xiao Zhu, Youyou Tao, Ruilin Zhu, Dezhi Wu, Wai-kit Ming. . . . 668

A Virtual Community of Practice to Support Physician Uptake of a Novel Abortion Practice: Mixed Methods Case Study (e34302)
Sheila Dunn, Sarah Munro, Courtney Devane, Edith Guilbert, Dahn Jeong, Eleni Stroulia, Judith Soon, Wendy Norman. . . . 684

Research on Digital Technology Use in Cardiology: Bibliometric Analysis (e36086)
Andy Yeung, Stefan Kulnik, Emil Parvanov, Anna Fassl, Fabian Eibensteiner, Sabine Völkl-Kernstock, Maria Kletecka-Pulker, Rik Crutzen, Johanna Gutenberg, Isabel Höppchen, Josef Niebauer, Jan Smeddinck, Harald Willschke, Atanas Atanasov. . . . 701

Effect of a Popular Web Drama Video Series on HIV and Other Sexually Transmitted Infection Testing Among Gay, Bisexual, and Other Men Who Have Sex With Men in Singapore: Community-Based, Pragmatic, Randomized Controlled Trial (e31401)
Rayner Tan, Wee Koh, Daniel Le, Sumita Banerjee, Martin Chio, Roy Chan, Christina Wong, Bee Tai, Mee Wong, Alex Cook, Mark Chen, Chen Wong. . . . 718

Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: 6-Month Follow-up Study of a Randomized Clinical Trial (e37480)
Laura Garcia, Brandon Birckhead, Parthasarathy Krishnamurthy, Ian Mackey, Josh Sackman, Vafi Salmasi, Robert Louis, Carina Castro, Roselani Maddox, Todd Maddox, Beth Darnall. . . . 732

Monitoring User Opinions and Side Effects on COVID-19 Vaccines in the Twittersphere: Infodemiology Study of Tweets (e35115)
Beatrice Portelli, Simone Scaboro, Roberto Tonino, Emmanuele Chersoni, Enrico Santus, Giuseppe Serra. . . . 764

Multimodal Chronic Pain Therapy for Adults via Smartphone: Randomized Controlled Clinical Trial (e36114)
Yolanda Morcillo-Muñoz, Antonio Sánchez-Guarnido, Silvia Calzón-Fernández, Isabel Baena-Parejo. . . . 781

Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation (e31810)
Malte Schmieding, Marvin Kopka, Konrad Schmidt, Sven Schulz-Niethammer, Felix Balzer, Markus Feufel. . . . 801

XSL•FO
RenderX

Commentary

Through the Narrative Looking Glass: Commentary on “Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review” (e38513)
Charlene Weir. . . . 148

Viewpoints

Improving Research Patient Data Repositories From a Health Data Industry Viewpoint (e32845)
Chunlei Tang, Jing Ma, Li Zhou, Joseph Plasek, Yuqing He, Yun Xiong, Yangyong Zhu, Yajun Huang, David Bates. . . . 151

Measuring Health-Related Quality of Life With Multimodal Data: Viewpoint (e35951)
Ieuan Clay, Francesca Cormack, Szymon Fedor, Luca Foschini, Giovanni Gentile, Chris van Hoof, Priya Kumar, Florian Lipsmeier, Akane Sano, Benjamin Smarr, Benjamin Vandendriessche, Valeria De Luca. . . . 155

Digital Transformation of Face-To-Face Focus Group Methodology: Engaging a Globally Dispersed Audience to Manage Institutional Change at the World Health Organization (e28911)
Gaya Gamhewage, Mohamed Mahmoud, Anna Tokar, Melissa Attias, Christos Mylonas, Sara Canna, Heini Utunen. . . . 169

Types of Errors Hiding in Google Scholar Data (e28354)
Romy Sauvayre. . . . 174

Letters to the Editor

Application of Nonfungible Tokens to Health Care. Comment on “Blockchain Technology Projects to Provide Telemedical Services: Systematic Review” (e34276)
John Gambril, Carter Boyd, Jamal Egbaria. . . . 750

Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on “Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial” (e34437)
Chih-Wei Chen, James Wei. . . . 752

Authors Reply to: Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on “Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial” (e37413)
Aya Adly, Mahmoud Adly, Afnan Adly. . . . 754

Learning More About the Effects of Gamification on Physical Activity. Comment on “Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials” (e36396)
Cheng-Hsien Hung, Yung-Shuo Kao. . . . 757

Authors’ Reply to: Learning More About the Effects of Gamification on Physical Activity. Comment on “Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials” (e38212)
Alexandre Mazeas, Martine Duclos, Bruno Pereira, Aïna Chalabaev. . . . 759

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data From the Health Information National Trends Survey (e39712)
Alexandra Greenberg-Worisek, Shaheen Kurani, Lila Finney Rutten, Kelly Blake, Richard Moser, Bradford Hesse. 761

XSL•FO
RenderX

Review

Regulatory Frameworks for Clinical Trial Data Sharing: Scoping Review

Nachiket Gudi[1], BSc, MPH; Prashanthi Kamath[1], BSc, MPH; Trishnika Chakraborty[1], MDS; Anil G Jacob[1], PhD; Shradha S Parsekar[2], BHMS, MPH; Suptendra Nath Sarbadhikari[1], MBBS, PhD; Oommen John[3,4], MBBS, MD

[1]The George Institute for Global Health, New Delhi, India

[2]Public Health Evidence South Asia, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, India

[3]The George Institute for Global Health, University of New South Wales, New Delhi, India

[4]Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, India

Corresponding Author:
Oommen John, MBBS, MD
The George Institute for Global Health, University of New South Wales
308-309, Third Floor, Elegance Tower, Plot No. 8, Jasola District Centre
New Delhi, 110025
India
Phone: 91 114158809193
Fax: 91 1141588090
Email: ojohn@georgeinstitute.org.in

Abstract

Background: Although well recognized for its scientific value, data sharing from clinical trials remains limited. Steps toward harmonization and standardization are increasing in various pockets of the global scientific community. This issue has gained salience during the COVID-19 pandemic. Even for agencies willing to share data, data exclusivity practices complicate matters; strict regulations by funders affect this even further. Finally, many low- and middle-income countries (LMICs) have weaker institutional mechanisms. This complex of factors hampers research and rapid response during public health emergencies. This drew our attention to the need for a review of the regulatory landscape governing clinical trial data sharing.

Objective: This review seeks to identify regulatory frameworks and policies that govern clinical trial data sharing and explore key elements of data-sharing mechanisms as outlined in existing regulatory documents. Following from, and based on, this empirical analysis of gaps in existing policy frameworks, we aimed to suggest focal areas for policy interventions on a systematic basis to facilitate clinical trial data sharing.

Methods: We followed the JBI scoping review approach. Our review covered electronic databases and relevant gray literature through a targeted web search. We included records (all publication types, except for conference abstracts) available in English that describe clinical trial data–sharing policies, guidelines, or standard operating procedures. Data extraction was performed independently by 2 authors, and findings were summarized using a narrative synthesis approach.

Results: We identified 4 articles and 13 policy documents; none originated from LMICs. Most (11/17, 65%) of the clinical trial agencies mandated a data-sharing agreement; 47% (8/17) of these policies required informed consent by trial participants; and 71% (12/17) outlined requirements for a data-sharing proposal review committee. Data-sharing policies have, a priori, milestone-based timelines when clinical trial data can be shared. We classify clinical trial agencies as following either controlled- or open-access data-sharing models. Incentives to promote data sharing and distinctions between mandated requirements and supportive requirements for informed consent during the data-sharing process remain gray areas, needing explication. To augment participant privacy and confidentiality, a neutral institutional mechanism to oversee dissemination of information from the appropriate data sets and more policy interventions led by LMICs to facilitate data sharing are strongly recommended.

Conclusions: Our review outlines the immediate need for developing a pragmatic data-sharing mechanism that aims to improve research and innovations as well as facilitate cross-border collaborations. Although a *one-policy-fits-all* approach would not account for regional and subnational legislation, we suggest that a focus on key elements of data-sharing mechanisms can be used to inform the development of flexible yet comprehensive data-sharing policies so that institutional mechanisms rather than disparate efforts guide data generation, which is the foundation of all scientific endeavor.

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e33591) doi:10.2196/33591

KEYWORDS

clinical trial; data sharing; policy; scoping review

Introduction

Background

Data sharing from clinical trials is a contested space; it has gained salience particularly during the COVID-19 pandemic. Clinical data are defined as "the data, results, information, discoveries, inventions, processes and methods (whether patentable or not) resulting from or developed by investigator or study personnel in the performance of the clinical trial, but excludes all personal information and medical records" [1]. Clinical trial data sharing is defined as "sharing of anonymized, patient-level clinical trial data through established platforms thereby enhancing transparency, thus maximizing value of research and creating opportunities for external researchers to reanalyze, synthesize, replicate, and build upon previous evidence" [2,3]. Sharing anonymized individual participant data (IPD) along with other trial-generated data can often pave ways for informed clinical decisions. In particular, the secondary analysis of such clinical trial data helps in building on the body of existing evidence by consolidating data across smaller, underpowered studies. It is one of several cost-effective measures for augmenting a body of evidence in resource-constrained settings and in health emergencies [4]. The COVID-19 pandemic is considered a booster for clinical data sharing because many researchers and working groups have strongly advocated it [5-8]. Ideally, clinical trial data sharing needs to be harmonized and standardized for the global scientific community. However, to align with the purpose of research, data from human participants should benefit others, and data sharing is one of the best ways to achieve this.

Clinical trial agencies have provided guidelines for regulating data sharing in clinical research. In particular, the International Committee of Medical Journal Editors [9], the UK-based Pragmatic Clinical Trials Unit [10], and the US-based National Institutes of Health (NIH) StrokeNet [11] have developed guidelines for efficiently sharing and accessing data. Clinical trial registries and scholarly publications expect biomedical researchers to provide statements on sharing data during various stages of clinical trials (eg, at the time of trial registration, after the planned interim analysis, at the midterm, and at the end of the clinical trial), as applicable according to the respective data-sharing guidelines. These data-sharing guidelines aim to safeguard the privacy of study participants when data are used by a researcher to build on existing evidence (secondary research) and thereby maximize benefits for the public [12]. According to the clinical trial registration policy of the International Committee of Medical Journal Editors, there are prerequisites for data oversight or the presence of an institutional ethics committee to abide by the Good Clinical Practice guideline as outlined by the International Council for Harmonisation [13]. These prerequisites need to be operationalized through gaining the informed consent of study participants to ensure the safety of the study participants, investigators [13,14], and those involved in clinical trials. Furthermore, data sharing from clinical research is generally governed by national regulatory agencies in their respective locations [14,15].

Where such guidance exists, regulatory policy documents provide guidance on data sharing and access to data by ensuring participant safety and ethical compliance [16]. The emerging conflicts between data-sharing practices and potential threats to the privacy and confidentiality of trial participants are significant challenges faced by investigators in complying with data-sharing principles. The review of data-sharing guidelines by Blasimme et al [17] and a stakeholder consultation of 1329 scientists [18] demonstrated that the importance of data sharing in medical science is not sufficiently recognized. Technical, legal, and ethical barriers hinder data sharing from clinical trials. Technical barriers include lack of standardization, limited researcher capacity to build high-quality data, and a lack of financial incentives for data sharing. Legal barriers such as intellectual property rights (IPR), data ownership, concerns of data provider and data user regarding mutual benefits, and explicit informed consent for data sharing are threats to data sharing [3,5,17,18]. Similar concerns were shared by the public in a high-income setting where widely shared data could be a risk for patient privacy and could give rise to discrimination and exploitation [19].

A Complicated Issue

The competing interests of stakeholders involved in clinical trials make data sharing a complicated issue owing to factors relating to investments and existing legal frameworks surrounding IPR. The resistance from for-profit pharmaceutical corporations is also understandable when they advocate data exclusivity, given their financial investments in conducting clinical trials. Most large multicenter clinical trials are funded by for-profit pharmaceutical corporations. Besides being data generators, these corporations are investors and risk takers, as well as intellectual conceptualizers of complex scientific information. Not surprisingly, such corporations have the incentive, control, and power to restrict data sharing. The agreement on Trade-Related Aspects of Intellectual Property Rights at the World Trade Organization (WTO) identifies transparency and availability of the latest information as being pivotal to trade and policy [20-22]. An updated list of IPR measures specific to a region and country limits the control exercised by multinational corporations [23]. Concerns over clinical trial data sharing follow IPR-related issues that arise with the sharing of undisclosed trial information—a practice often referred to as "secret trial data" [24,25]. For-profit pharmaceutical corporations often resist, or lack interest in, data-sharing efforts through their data-exclusivity practices [26,27]. This can have, and has had, a negative impact on access to medicines and biologicals, including vaccines in low- and middle-income countries (LMICs) [28].

XSL•FO
RenderX

Institutional frameworks based on jurisdictions vary considerably. Data sharing from cross-border or multicountry and multisite randomized controlled clinical trials are generally not governed by a single (or even comparable) national legislation [29,30]. Although international regulations on data sharing are lacking, there are a few guidelines. In a multicenter and multicountry trial, there are context-specific issues such as the cost of trial completion and data sharing, subcontracting, and the use of third parties to complete the trial. Nuances in subcontracting the conduct of the trial to for-profit and not-for-profit organizations further make data sharing difficult. There are ethical complexities as well: large sample sizes are often possible only by recruiting participants from low-income countries owing to their larger populations and (often) poor clinical trial oversight and regulatory mechanisms [19,31]. Not surprisingly, therefore, the wider acceptance of data-sharing practices in the absence of a mandate to share clinical trial data has created uncertainties among clinical trial investigators [32,33]. With the paucity of surveys or academic syntheses to offer guidance on data sharing, it is necessary to collate evidence and classify this information to facilitate syntheses and comparability with regard to data-sharing practices.

Given the limitations in the existing landscape of clinical trials regarding data sharing, it must be noted that at the institutional level, systematic steps are being taken to shift data sharing in more institutionalized directions, which is laudable. This is based on disclosures by funders; nevertheless, such disclosures vary in degrees. For ease of comprehension, we view these in the binary categories of open- or controlled-access models of sharing data. Pursuant to this, trial investigators are implementing data sharing according to varied milestones, depending on the progress of the clinical trial. Thus, at the aggregate level, timelines for disclosures also vary. In brief, such principal investigators connect the level of data disclosure to the completion of varied milestones. Our review appropriately classifies this information.

Considering the aforementioned gaps, this review attempts to synthesize the existing state of practices around clinical trial data sharing. Our viewpoint is decidedly from a public health perspective because we believe that data sharing needs to be promoted for the public good. With this intention, we conducted a scoping review of the literature with the following objectives: to identify regulatory documents that have guided clinical trial investigators in trial data sharing and to explore the key elements of data-sharing mechanisms in these regulatory documents.

Methods

Overview

A scoping review approach facilitates an understanding of emerging evidence and is often considered the first step in research evidence development [34]. We followed the JBI methodology, as proposed in the methodological framework of Arksey and O'Malley [35] for scoping studies and the work on advancing this methodology by Levac et al [36]. The review protocol was developed a priori; however, because of the time-bound nature of this review, we could neither register nor publish the protocol. The JBI methodology has outlined six steps for the conduct of a scoping review: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data, (5) collating, summarizing, and reporting the results; and (6) stakeholder consultations [34-36]. These steps are further described in the following sections. The scoping review is reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [37].

Step 1: Identifying the Research Question

As we intended to synthesize a fast-growing but fragmented body of literature on regulatory documents for data sharing, we did not follow the typical *Population, Intervention, Comparison, and Outcomes* or *Population, Concept, and Context* approach to guide our article selection process because this topic falls beyond the scope of these and other review classifications [38]. We developed the objectives of this scoping review using an iterative approach. Stakeholders (a group of domain experts) from the data-sharing working group of the COVID-19 Clinical Research Coalition were involved in providing feedback on the objectives [39]. In particular, one of the authors (OJ) is a member of the data-sharing working group of the COVID-19 Clinical Research Coalition, and this review was undertaken as a specific deliverable with technical support from the data-sharing working group.

Step 2: Identifying Relevant Studies

We followed a 2-pronged approach for identifying regulatory documents: (1) literature search in scientific journals and (2) gray literature search. We conducted searches on MEDLINE (PubMed), SCOPUS, CINAHL (EBSCO), EMBASE, ProQuest, and Google Scholar using the keywords *data sharing policy*, *data sharing guidance*, *clinical trial data sharing*, and *individual participant data sharing*. The search was carried out by 2 authors (NG and SSP) on May 11, 2021. A detailed search strategy for each database is presented in Multimedia Appendix 1.

An alert was created between May 11, 2021, and August 31, 2021, on the aforementioned databases for the search strategy to further include articles as and when published. This step was deemed necessary because the topic is dynamic and published scholarly evidence has been emerging regularly since the COVID-19 pandemic began. The gray literature is an important source for gathering further evidence on data sharing. To populate a comprehensive list of trial agencies, we manually looked at the trial websites through a Google search. We also searched for data-sharing policies on the clinical trial agency websites (Multimedia Appendix 2 [16,40-64]). We further conducted reference screening of articles included at the full-text stage to identify any potential inclusions. All search results were uploaded into EndNote software (Clarivate), and duplicates were removed.

Step 3: Study Selection

The selection of studies was carried out by 2 authors (PK and TC) independently in 2 sequential stages, namely, title-abstract and full-text stages. We used this 2-stage strategy because the evidence suggests that there is no difference between the titles-first and title-abstract–together approaches [65]. Conflicts

XSL•FO
RenderX

on study selection were discussed until consensus was reached, or a senior team member (NG or OJ) acted as an arbitrator to decide on final inclusion or exclusion of the record. To ensure transparency in the study selection process, the number of records included at each stage was represented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 chart, along with the reasons for exclusion of studies during the full-text screening [66]. The study selection was based on the following inclusion and exclusion criteria:

- Inclusion criteria: as this is a broad and emerging topic, we did not limit our review to studies with any specific research designs. We included studies that described any clinical trial data–sharing policies or standard operating procedures, which are defined as detailed, written instructions to achieve uniformity of the performance of a specific function upholding the goal of the Good Clinical Practice guideline [67]. This may include but not be limited to evidence synthesis papers such as systematic reviews and rapid evidence synthesis. We also included commentaries, editorials, and policy briefs; however, we restricted our search to studies published in English.
- Exclusion criteria: we excluded conference abstracts because these are susceptible to changes (eg, content or title) at the completed-manuscript stage, making it challenging to locate them. As our search was comprehensive, our decision to exclude conference abstracts would have had minimal impact in terms of the number of articles retrieved at the title-abstract stage.

Step 4: Charting the Data

Data coding was carried out independently by 2 authors (PK and TC) using a predesigned yet flexible data-coding template (Multimedia Appendix 3). The study team members were consulted before finalizing the data-coding template, and minor modifications were incorporated based on the team's feedback. The data-coding template was pretested as suggested by Levac et al [36] and Daudt et al [68]. Extracted data were coded as per characteristics of the regulatory documents (name of trial agency, type of regulatory document, recent version, regulatory document and policy scope, country of origin, geographical scope, scientific scope, timeline, and grant limit) and data-sharing mechanism (need for data-sharing agreement, informed consent, type of review committee, timeline to share and access data, cost of data sharing, and data-sharing model). Clinical trial agencies were categorized as being either *for-profit pharmaceutical trial agencies*, *federal or national regulatory agencies* (publicly funded trial agencies), *academic institutions* (affiliated to universities), *nonprofit research organizations*, or *networks/consortia of clinical trial units*.

Depending on the description of the data access requirement, we coded the data access model as either an *open-access model* or a *controlled-access model*. If the anonymized trial data are made available to the public without submission of a proposal or without an approval process and no limitations or restrictions on data use, it was coded as an *open-access* model. If the data request is reviewed against prespecified criteria by internal or external review committees, we coded it as a *controlled-access* model [69]. Furthermore, if the information in the included documents was insufficient, we referred to the source (often websites) of the respective documents to elicit further information; there was 1 such policy document [40,70].

Step 5: Collating, Summarizing, and Reporting the Results

We used a narrative synthesis approach to summarize findings to provide a comprehensive view of an emerging topic owing to its potentially large volume and heterogeneous nature. Critical appraisals of studies and summarizing results from individual studies were beyond the scope of this review.

Step 6: Stakeholder Consultation

As we sought to code data from the available pool of scientific guidelines, global standards, and national policies, conducting a stakeholder consultation for summarizing the results was beyond the scope of this project. Stakeholders from the data-sharing working group of the COVID-19 Clinical Research Coalition were consulted for finalizing the review objectives as described in step 1.

Results

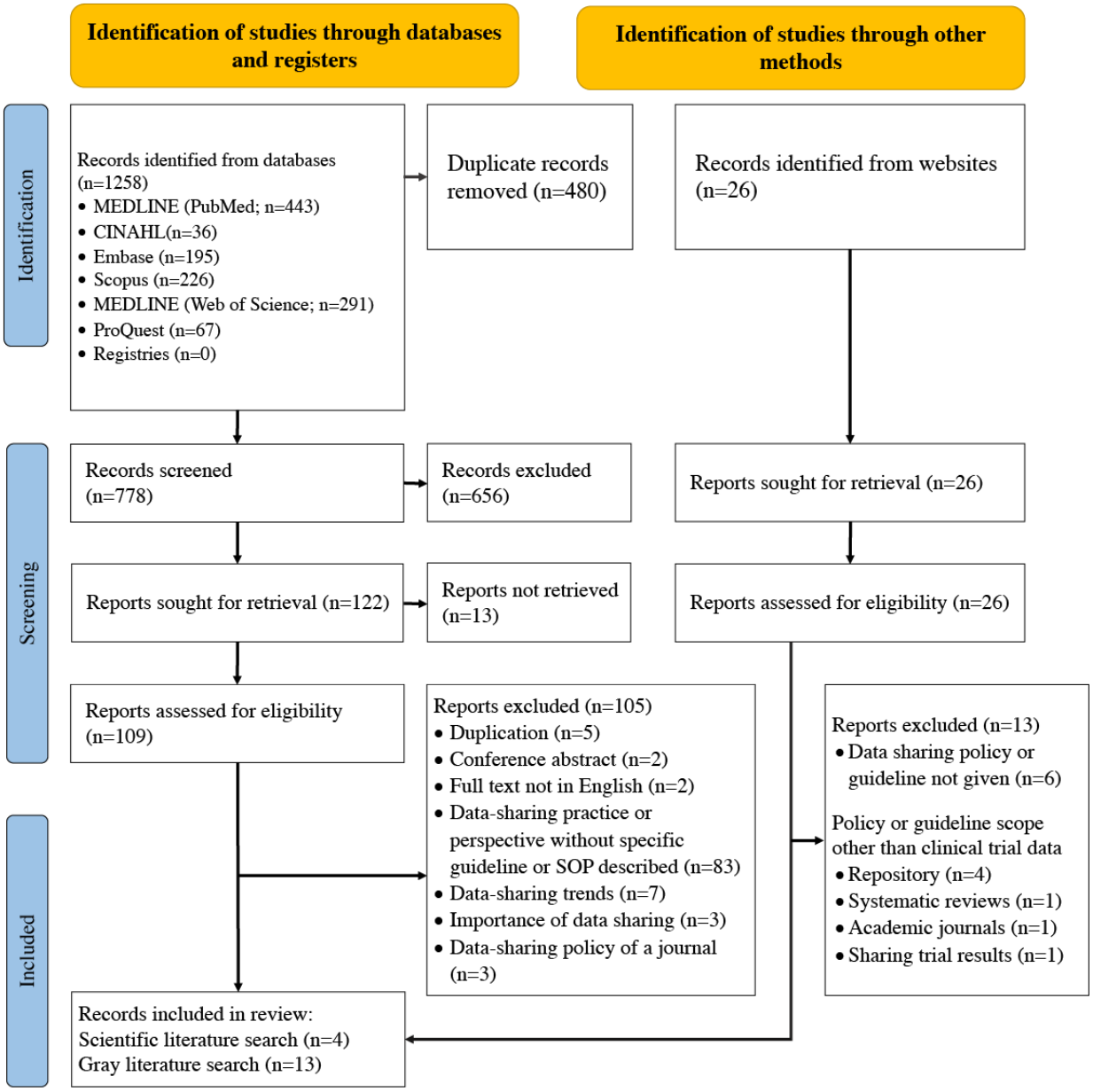
We have presented the findings of this scoping review based on the evidence identified through a scientific literature search and a gray literature search.

Findings From Scientific Literature Search

The initial search yielded a total of 1258 records; after 480 (38.16%) duplicates were removed, we screened 778 (61.84%) titles and abstracts, and included 109 (8.66%) reports for the eligibility check. After the full-text screening of these 109 articles, we finally included 4 (3.7%) for the data coding [70-73]. Papers with a focus on trends in data sharing and importance of data sharing were excluded. The PRISMA 2020 chart is presented in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 chart. SOP: standard operating procedure.



Findings From Gray Literature Search

We found few academic publications on regulatory frameworks from clinical research. Therefore, we decided to conduct a gray literature search. During the full-text review stage, we made a note of the referred clinical trial agencies, and the websites of these agencies were examined for clinical trial data–sharing policies—either a policy, a standard operating procedure, or a guiding document on data sharing. We included a total of 13 [10,11,16,41-44,74-79] trial agencies that had at least one data-sharing policy or guiding document. A list of trial agencies can be found in Table 1. We have provided the list of trial websites we examined, as well as those excluded, in Multimedia Appendix 2.

XSL•FO
RenderX

Table 1. Summary of articles included in scientific journals.

Study	Name of clinical trial agency	Type of clinical trial agency	Type of regulatory document	Geographical scope	Scientific scope
Nisen and Rockhold, 2013 [72]	GSK[a]-sponsored trials	For-profit pharmaceutical trial agency	Policy	Global	GSK-sponsored trials
Ross et al, 2018 [70]	YODA[b]	Academic	Policy	Not reported	YODA partners (Medtronic and Johnson & Johnson)
Pencina et al, 2016 [73]	SOAR[c] initiative	For-profit pharmaceutical-academia collaboration	Policy	United States	BMS[d]-sponsored phase I to phase IV trials (trials completed after January 2008)
Mitka, 2015 [71]	IOM[e]	Nonprofit regulatory agencies	Guideline	Not reported	Not reported

[a]GSK: GlaxoSmithKline.

[b]YODA: Yale Open Data Access.

[c]SOAR: Supporting Open Access for Researchers.

[d]BMS: Bristol Myers Squibb.

[e]IOM: Institute of Medicine.

Summary of Articles Included From Scientific Literature Search

We included perspective papers that summarized specific regulatory documents. The scientific scope of these regulatory documents varied significantly. Of the 4 documents included, 3 (75%) [70,72,73] referred to a specific policy document, whereas 1 (25%) [71] referred to a data-sharing guideline. These documents were published between 2013 and 2018. The trial agencies specified in these documents were a for-profit pharmaceutical firm (GlaxoSmithKline [GSK]-sponsored clinical trials) [72], an academic institution (Yale Open Data Access) [70], a pharmaceutical-academia collaboration (Supporting Open Access for Researchers) [73], and a nonprofit regulatory agency (Institute of Medicine) [71]. The GSK policy is applicable to GSK-sponsored trials globally, whereas the Supporting Open Access for Researchers initiative is applicable to Bristol Myers Squibb (BMS)–sponsored phase I-IV interventional clinical trials completed after 2008 in the United States. The remaining 50% (2/4) of the documents [70,71] did not specify geographical scope. The Yale Open Data Access project policy is applicable to their partner clinical trial agencies such as trial data sponsored by Medtronic and Johnson & Johnson clinical trial data [70] (Table 1).

Summary of Regulatory Documents Identified Through Gray Literature Search

We report the summary of the included regulatory documents in the gray literature in Table 2. In case of multiple versions, the most recent version of these documents was considered for coding the data. All the clinical trial agencies were from the United States and the United Kingdom. For-profit pharmaceutical trial agencies [41,74] (Celgene, Pharmaceutical Research and Manufacturers of America, and European Federation of Pharmaceutical Industries and Associations) follow regulations applicable to the trials conducted in the United States and the European Union. The NIH [11,16,42-44,75] and the Medical Research Council [76] have a data-sharing policy applicable to their own funded clinical trials. The policies of NIH-affiliated agencies [16,42] are applicable to trials with a grant limit of US $500,000 or more in direct costs in any year of the proposed project period. None of the other policy documents specify a grant limit for their applicability (Table 2).

XSL•FO
RenderX

Table 2. Summary of regulatory documents (gray literature).

Policy	Recent policy version	Country of origin	Geographical scope	Scientific scope	Timeline	Grant limit (US $)
Celgene [41]	Version 4, 2017	United States	United States and European Union	Compound and indication trials	__[a]	—
PhRMA[b] and EFPIA[c] [74]	2013	United States	United States and European Union	—	—	—
NIH[d]-1 [42]	2003	United States	—	NIH-funded studies	From 2003 to 2023	≥500,000 or more (in direct costs[e])
NIH-2 [16]	2020	United States	—	NIH-funded or NIH-conducted research	From 2023	≥500,000 or more (in direct costs[e])
NIH-NHLBI[f] [43]	2014	United States	—	NHLBI-funded studies (related to heart, lung, and blood–related research)	From 2003 to 2023	—
NIH-NCI[g] [44]	2016	United States	—	NCI-supported Cancer Moonshot studies	On or after 2017	—
MRC[h] [76]	2015	United Kingdom	—	Publicly funded clinical trials units	—	—
NIHR[i] [75]	Version 1, 2019	United Kingdom	—	NIHR-funded research studies	—	—
EMA[j] [79]	Policy 0070, 2018	United Kingdom	—	For academic and noncommercial research purposes	Data held after July 1, 2015	—
NIH StrokeNet [11]	Version 1, 2014	United States	—	Clinical trials conducted in the NIH StrokeNet network	—	—
PCTU[k] [10]	Version 5, 2019	United Kingdom	—	Clinical research data held on PCTU servers	From January 1, 2014	—
PCORI[l] [77]	2018	United States	—	Research projects funded by PCORI	—	—
UKCRC[m] [78]	2021	United Kingdom	—	UKCRC-registered clinical trials unit network	—	—

[a]Not mentioned.

[b]PhRMA: Pharmaceutical Research and Manufacturers of America.

[c]EFPIA: European Federation of Pharmaceutical Industries and Associations.

[d]NIH: National Institutes of Health.

[e]In any year of the proposed project period through grants, cooperative agreements, or contracts.

[f]NHLBI: National Heart, Lung, and Blood Institute.

[g]NCI: National Cancer Institute.

[h]MRC: Medical Research Council.

[i]NIHR: National Institute for Health Research.

[j]EMA: European Medicines Agency.

[k]PCTU: Pragmatic Clinical Trials Unit.

[l]PCORI: Patient-Centered Outcomes Research Institute.

[m]UKCRC: UK Clinical Research Collaboration.

Data-Sharing Mechanisms

We reviewed data-sharing mechanisms from the documents and summarize the key specifications in Table 3. Most (11/17, 65%) [10,40-42,70-73,75-78] of the clinical trial agencies require a data-sharing agreement between the data requester and the clinical trial agency. The specific requirements to access data and obligations with good data-sharing principles are highlighted in the data-sharing agreement. We could access data-sharing–agreement templates from 27% (3/11) [10,42,76] of these trial agencies. At these trial agencies, data-sharing agreements between the data requester and clinical trial agency are mutual and nondisclosable in nature. Data-sharing agreements ensure that appropriate data anonymization procedures are followed, thereby minimizing the chance of identifying the study participant. These agreements further prohibit data users from sharing data with third parties. However, the legal actions in cases of noncompliance are not clearly defined. The data request process is facilitated either through a website registration procedure or submission of a data request form. Basic information about the principal investigator of secondary research, key personnel, and the research proposal, which includes project title, scientific abstract, brief project

XSL•FO
RenderX

background and statement of project significance, specific aims, research methods, narrative summary, project timeline, dissemination plan, and bibliography, are standard requirements across various trial registries. Data requesters need to mention their proposed process for data management and the process of making the resulting publications available to the public. Not all the trial agencies require an independent review committee. However, an independent review committee would possess the right to decide on sharing the data.

Nearly half (8/17, 47%) [10,40,41,43,44,70,72,75,76] of the trial agencies mentioned that informed consent for data sharing should be included in the broader informed consent for research. However, curiously, none of the policy documents mention that consent is mandated. Of these 8 policy documents, 2 (25%) [40,44,70] make statements to the effect that data-sharing practices should follow the data-sharing statement presented in the broader informed consent.

Of the 17 clinical trial agencies, only 5 (29%) [10,40,70-72,76] specified a general timeline within which data would be shared. Of these 5 agencies, 2 (40%) specified that data would be shared along with the publication, 1 (20%) [76] specified a timeline of 18 months of trial completion, and 1 (20%) [10] specified a timeline of 24 months of trial completion. Of these 5 agencies, 1 (20%) [71] specified a separate timeline for data underlying the results, that is, no later than 6 months from the time of publication, and no later than 18 months from the time of publication for IPD. The time limit to access data was described by 29% (5/17) of the clinical trial agencies: 12 months by 80% (4/5) [10,40,70,72] of these agencies and 7 years by 20% (1/5) [77]. Of the 17 clinical trial agencies, 1 (6%) [75] stated that data access would be granted as per the agreement.

Of the 17 trial agencies, 12 (71%) [10,40-44,70,72,74,75,77-79] provided a sufficient description of their data access model. Of these 12 trial agencies, 10 (83%) [10,40-43,70,72,74,75,77,78] practiced a controlled-access model, whereas 2 (17%) [44,79] followed both open- and controlled-access models based on the type of data. Only 35% (6/17) [16,40,42,43,70,71,76] of the clinical trial agencies specified who is to bear the cost of data sharing. Policies identify varied sources responsible for the cost of data sharing: independent funder [40,70], trial sponsor [76], or the clinical trial agency itself [16,42,43]. The Institute of Medicine, a clinical trial agency, has stated that the cost of data sharing should be shared by the clinical trial sponsor and the secondary data user [71]. Clinical trial agencies such as the NIH [16,42], the NIH-affiliated trial agencies [43], and the Medical Research Council [76] encourage trial investigators to estimate data-sharing expenses in the grant application. Most (14/17, 82%) of the data-sharing policies were applicable to IPD sharing. The other clinical trial data in Table 3 refer to case report forms, protocols, reporting, and analysis plans. The included data-sharing policies varied widely in terms of the terminology used to describe the *type of trial data*. To standardize our interpretation of clinical trial data applicability on the type of clinical trial data, we referred to the definitions given in the policy documents. For example, the NIH 2003 policy applies to underlying research data of the final summary statistics and results [42]. By contrast, the new NIH 2023 policy is applicable to all the scientific factual data that are accepted in the scientific community to validate and replicate research findings [16] (Table 3).

XSL•FO
RenderX

Table 3. Key elements of data-sharing mechanisms in regulatory documents.

	Requires data-sharing agreement	Requires review committee	Requires informed consent	Specifies timeline to share data	Specifies time limit to access data	Specifies data-sharing model	Specification on cost of data sharing	Requires sharing of IPD[a]	Specification on sharing of other clinical trial data
GSK[b]-sponsored trials [72]	Yes	Yes	Yes (applicable from 2013)	No	Yes	Yes	No	Yes	Yes
YODA[c] [70]	Yes	Yes	Yes (applicable from 2014)	No	Yes	Yes	Yes	Yes	Yes
SOAR[d] initiative	Yes	Yes	No	No	No	Cannot ascertain	No	Yes	Yes
IOM[e] [71]	Yes	Yes	No	Yes	No	Cannot ascertain	Yes	Yes	Yes
Celgene [41]	Yes	Yes	Yes (applicable from 2014)	No	No	Yes	No	Yes	Yes
PhRMA[f] and EFPIA[g] [74]	No	Yes	No	No	No	Yes	No	Yes	Yes
NIH[h]-1 [42]	Yes	Yes	No	No	No	Yes	Yes	No	Yes
NIH-2 [16]	No	No	No	No	No	Cannot ascertain	Yes	Yes	Yes
NIH-NHLBI[i] [43]	No	No	Yes	Yes	No	Yes	Yes	Yes	No
NIH-NCI[j] [44]	No	No	Yes (if conducting research would benefit public health)	Yes	No	Yes	No	Yes	No
MRC[k] [76]	Yes	Yes	Yes	Yes	No	Cannot ascertain	Yes	Yes	No
NIHR[l] [75]	Yes	Yes	Yes	No	Yes	Yes	No	Yes	Yes
EMA[m] [79]	No	Yes	No	No	No	Yes	No	Yes	Yes
NIH StrokeNet [11]	No	No	No	No	No	Cannot ascertain	No	No	No
PCTU[n] [10]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes
PCORI[o] [77]	Yes	Yes	No	No	Yes	Yes	No	Yes	No
UKCRC[p] [78]	Yes	No	No	No	No	Yes	No	No	No
Total, n (%; yes or no)	11 (65)	12 (71)	8 (47)	5 (29)	5 (29)	12 (71)	6 (35)	14 (82)	11 (65)

[a]IPD: individual participant data.

[b]GSK: GlaxoSmithKline.

[c]YODA: Yale Open Data Access.

[d]SOAR: Supporting Open Access for Researchers.

[e]IOM: Institute of Medicine.

[f]PhRMA: Pharmaceutical Research and Manufacturers of America.

[g]EFPIA: European Federation of Pharmaceutical Industries and Associations.

[h]NIH: National Institutes of Health.

[i]NHLBI: National Heart, Lung, and Blood Institute.

[j]NCI: National Cancer Institute.

[k]MRC: Medical Research Council.

XSL•FO
RenderX

[l]NIHR: National Institute for Health Research.
[m]EMA: European Medicines Agency.
[n]PCTU: Pragmatic Clinical Trials Unit.
[o]PCORI: Patient-Centered Outcomes Research Institute.
[p]UKCRC: UK Clinical Research Collaboration.

Discussion

Principal Findings

Data-sharing practices have not been a characteristic of most randomized controlled clinical trials. There may be many reasons for this: for example, incentives to shift the status quo away from proprietary models of holding on to data have often remained diminutive. For producers of data, such as large corporations, the alignment of financial investments dovetails well with the desire to amortize these costs in terms of limiting data sharing to the originators of the data. On the demand side, firms or entities with the technical capacities to use such data are limited to competitor firms in the mostly capitalist Organisation for Economic Co-operation and Development countries. Emerging market firms from Brazil, Russia, India, China, and South Africa, or other nations with relatively more sophisticated technical capacities in reverse engineering are another class of potential consumers of such data. Institutionally, at the international level, the WTO has governed international trade in goods and services since its inception in 1995. As global trade requires abiding by WTO standards, there has been a cross-national harmonizing effect. This affects all goods and services in international trade. Global standards for goods and services have to be followed by all; to a large degree, this overrides national considerations, and therefore data-exclusivity practices are also introduced to harmonize transfers of both goods and the generation of data over the course of trade in services as well. As far as data-exclusivity practices are concerned, these are buttressed by domestic legal frameworks in the wealthy countries that generally initiate such trials. In the early 2000s, the Doha Declaration as well as debates regarding compulsory licensing episodically brought such issues to the fore but without abiding institutional shifts [28].

However, as the COVID-19 pandemic progressed, the lack of robust evidence hampered effective therapeutic and public health interventions, resulting in widespread panic as cases surged. Despite a large number of clinical trials aimed at repurposing existing interventions for managing COVID-19 and several promising drug candidate interventions undergoing clinical trials, the scientific community was unable to collaborate and synergize efforts. It is possible that this was on account of regulatory and policy bottlenecks that hampered clinical trial data sharing. This scoping review was intended to identify regulatory frameworks and policy guidance that support clinical trial data sharing. We included regulatory documents (n=4) from the scientific literature search as well as from the gray literature search (n=13). Our results indicate that clinical trial regulatory frameworks aim to make data available for noncommercial use for researchers. We noted that clinical trial agencies lack comprehensive approaches that facilitate data sharing. Of the 17 regulatory documents reviewed, 11 (65%) mandate the need for a data-sharing agreement, 8 (47%) require informed consent, 12 (71%) mandate a proposal review committee to oversee the data sharing, and 5 (29%) specify timelines for data sharing and a time limit to access data. A significant proportion (12/17, 71%) of these documents describe different data-sharing models: 83% (10/12) were related to IPD sharing and 33% (4/12) provided specifications regarding the cost of sharing data.

Data sharing is widely advocated as a norm in clinical research. However, regulatory frameworks and policy guidance to support researchers and institutions to share clinical trial data continue to lag behind such norms. This gap between the intention to share data and the lack of supportive regulatory and policy frameworks can be attributed to the direct or indirect effects connected with data sharing. At a macrolevel, this could be due to issues pertaining to IPR, differences across regulatory guidelines in high-income countries and LMICs, the variations in commercial interests of funding sources, and the potentially high economic benefits from data exclusivity [80-82]. Besides these issues, researchers are concerned about ensuring the privacy and confidentiality of study participants; although the informed consent procedures have provisions to seek participants' consent for data sharing and secondary use of their data, these are rarely implemented in practice. In practice, institutional ethics review boards often resort to myopic approaches when approving clinical trials that propose broader informed consent for data sharing and secondary use of data. These could also hamper the efforts of clinical trial investigators seeking to incorporate specific data-sharing clauses in the informed consent procedures. Moreover, the cost associated with data sharing, potential threats to confidentiality, academic credit, and investigator capacity to standardize data in a shareable manner are some concerns from a researcher's perspective [80-82]. Our gray literature screening of clinical trial websites showed that not every clinical trial agency is guided by its own data-sharing regulatory document (Multimedia Appendix 2). We noted that policies for clinical trial data sharing are evolving. For instance, the UK Clinical Research Collaboration published a standard operating procedure in 2021 to share participant data [78]; however, it is still at the development stage. Similarly, GSK recently agreed to share deidentified participant data [72].

If present, robust data-sharing practices often come into play when there is a legal agreement between the data requester and the data-sharing agency. The regulatory areas involved in the data-sharing mechanism guide investigators to share data in an appropriate manner to protect participant autonomy and data confidentiality. However, a formal agreement between a data requester and the trial agency is often not outlined in the data-sharing regulatory documents identified in this review. Existing regulatory and policy documents suggest that open access to clinical trial data may not be reliable because of higher chances of fraudulent reports or erroneous secondary analyses.

XSL•FO
RenderX

Given that clinical trial participants are often from multiple sites across the world, anonymization practices for the data must meet the respective countries' regulatory requirements. In the policies we reviewed, the details of cost of data sharing for the infrastructure and maintenance, data standardization, harmonization of data, and data quality assurance have not been described, let alone specified. Up-front resource investments for building sustainable and comprehensive data-sharing platforms with standardized data elements and user-friendly interfaces are likely to enhance the quality, accessibility, and usability of shared data. However, these may be expensive and financially untenable in LMICs. Of the 17 policies, only 6 (35%) [16,40,42,43,70,71,76] mentioned the cost of data sharing. Core clinical trial sponsors and agencies such as the NIH have recognized the need for supporting investigators for data sharing. The NIH states that in grant applications, data-sharing expenses can be estimated separately [16,42]. Such cost sharing is identified as one of the sustainable ways to achieve data sharing and was advocated at a public workshop conducted by the National Academies of Sciences, Engineering, and Medicine in November 2019 [32]. The Institute of Medicine is the only entity with a policy that highlighted the sharing of cost by the sponsor and data user [71].

Of the 17 clinical trial agencies, 5 (29%) [10,43,44,71,76] specified timelines to share data and these specifications referred to trial completion [10,71,76] or publication [43,44] as the milestones for sharing data appropriately. The benefits of data sharing can be maximized when it is encouraged at almost all stages of clinical trials; however, specifications regarding data sharing across major stages of clinical trials are often missing [83]. Furthermore, we found that none of the regulatory documents specified incentives or any kind of reward for data sharing. In addition, specifications relating to noncompliance with regard to data sharing are not clear. Data sharing from large multicenter international clinical trials is challenging because of the varied practices as guided by the respective national regulatory bodies [84]. We found that informed consent and legal agreements for data access are not a requirement for all policies. The regulatory frameworks do not cover all key elements of the data-sharing mechanisms discussed in this scoping review. However, it is encouraging to see the scope of these policies covering IPD sharing, rather than limiting sharing to overall clinical trial data.

The goal of any research involving human participants is to improve the health and quality of life of humans. Therefore, it is the need of the hour to look at data sharing with a moral and ethical lens for the public good. It is important to weigh the risks against the benefits of data sharing and find ways to overcome or mitigate the risks. Any data-sharing attempt without considering the trial funders is unlikely to be successful. Data sharing requires collaboration among apex federal or national trial agencies, academic institutions, and profit-based pharmaceutical clinical trial agencies. The cost of sharing data is another unexplored area that needs to be addressed. Either the main trial agency or another trial funder can play an important role in providing financial assistance and the capacity building of investigators for data sharing. A comprehensive data-sharing policy may not be feasible, given the diverse approaches in clinical research, geographies, and the scientific merit of a given clinical trial. Regulatory frameworks need to acknowledge these factors when standardizing data-sharing processes and provide a clear description for trial investigators rather than a broader document in support of data sharing. There is a strong need to define policy scope in terms of type of clinical trial, type of clinical trial data, and single-center or multicenter (including multicountry) trials, as well as specifications for privately funded and publicly funded trials or commercial and noncommercial funders. A clear distinction between *mandated requirements* and *supportive requirements* during the data-sharing process would help investigators to practice data sharing. To provide better participant privacy and confidentiality, there should be a neutral party to check the information in the data set. Creating an independent review panel to decide on the accessibility of the clinical data should be encouraged. A systematic review on increasing data sharing in health and medical research showed that there is a lack of research on evidence-based data sharing [85]. Nurturing clinical trial investigators, clinical trial funders, and academicians with rewards for data sharing should be encouraged. Data sharing from clinical trials is a daunting task; nevertheless, it is important to make this process easy for researchers, university academics, clinical trial agencies, and funders considering the benefits of data sharing. Ensuring viable, efficient, and feasible data-sharing mechanisms tailored to stakeholders' interests and bound to ethical and legal concerns is the way forward.

Strengths and Limitations

This scoping review used a systematic, replicable, and rigorous approach to summarize evidence. By using well-defined search terms, database searches, and screening of articles, our processes were rigorous because these were carried out independently by 2 authors (NG and SP or PK and TC). We carried out an extensive gray literature search and reference screening of the articles that were included at the full-text screening stage. Although stakeholders were contacted while finalizing the objectives, we were unable to consult them while drafting the results because of the nature of the review, and we would like to acknowledge this as a limitation. We had initially planned to carry out searches in scientific databases but later decided to perform the gray literature search, and this is reported as a deviation from our protocol.

Conclusions

This scoping review used a rigorous methodology to support clinical trial data sharing. Standardizing data-sharing processes by framing a more focused and concise policy with key elements of data-sharing mechanisms could be feasible and easier to practice than a single, rigid, and comprehensive data-sharing policy. We believe that to uncover the complexities and make data sharing a reality for the public good, negotiations around stakeholder interests are crucial. During and after the COVID-19 pandemic, and to paraphrase what Victor Hugo once remarked in another context, clinical data sharing may well be "an idea whose time has come."

XSL•FO
RenderX

Acknowledgments

This work was supported by the World Health Organization and TDR, the Special Programme for Research and Training in Tropical Diseases (P20-00060). This project was undertaken under the auspices of the data-sharing working group of the COVID-19 Clinical Research Coalition.

Authors' Contributions

OJ, NG, and the data-sharing working group of the COVID-19 Clinical Research Coalition were responsible for the conceptualization of this study. NG, OJ, PK, and SSP were responsible for the methodology. PK, TC, NG, and SSP were responsible for data curation. The formal analysis was performed by NG, PK, TC, and OJ. The project administration was the responsibility of NG with supervision by OJ. NG, SP, PK, TC were in charge of the software. The original draft was written by NG, PK, OJ, TC, and SNS. AGJ provided critical feedback on the manuscript. NG, PK, AGJ, TC, OJ, SSP, and SNS were responsible for writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 22 KB - jmir_v24i5e33591_app1.docx]

Multimedia Appendix 2
Gray literature search.
[DOCX File , 29 KB - jmir_v24i5e33591_app2.docx]

Multimedia Appendix 3
Data-coding template.
[DOCX File , 24 KB - jmir_v24i5e33591_app3.docx]

References

1. Clinical Trial Data definition. Law Insider. URL: https://www.lawinsider.com/dictionary/clinical-trial-data [accessed 2021-09-12]
2. Kochhar S, Knoppers B, Gamble C, Chant A, Koplan J, Humphreys GS. Clinical trial data sharing: here's the challenge. BMJ Open 2019 Aug 21;9(8):e032334 [FREE Full text] [doi: 10.1136/bmjopen-2019-032334] [Medline: 31439612]
3. Naudet F, Siebert M, Pellen C, Gaba J, Axfors C, Cristea I, et al. Medical journal requirements for clinical trial data sharing: ripe for improvement. PLoS Med 2021 Oct;18(10):e1003844 [FREE Full text] [doi: 10.1371/journal.pmed.1003844] [Medline: 34695113]
4. Ohmann C, Banzi R, Canham S, Battaglia S, Matei M, Ariyo C, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. BMJ Open 2017 Dec 14;7(12):e018647 [FREE Full text] [doi: 10.1136/bmjopen-2017-018647] [Medline: 29247106]
5. Fegan G, Cheah P, Data Sharing Working Group. Electronic address: WorkingGroups@covid19crc.org. Solutions to COVID-19 data sharing. Lancet Digit Health 2021 Jan;3(1):e6 [FREE Full text] [doi: 10.1016/S2589-7500(20)30273-9] [Medline: 33735070]
6. Harrison PW, Lopez R, Rahman N, Allen SG, Aslam R, Buso N, et al. The COVID-19 Data Portal: accelerating SARS-CoV-2 and COVID-19 research through rapid open access data sharing. Nucleic Acids Res 2021 Jul 02;49(W1):W619-W623 [FREE Full text] [doi: 10.1093/nar/gkab417] [Medline: 34048576]
7. Li R, von Isenburg M, Levenstein M, Neumann S, Wood J, Sim I. COVID-19 trials: declarations of data sharing intentions at trial registration and at publication. Trials 2021 Feb 18;22(1):153 [FREE Full text] [doi: 10.1186/s13063-021-05104-z] [Medline: 33602275]
8. Call for a global health data sharing framework for global health emergencies. The InterAcademy Partnership (IAP) Science Health Policy. 2022. URL: https://www.interacademies.org/publication/call-global-health-data-sharing-framework-global-health-emergencies [accessed 2022-03-15]
9. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen J, et al. Sharing clinical trial data — a proposal from the international committee of medical journal editors. N Engl J Med 2016 Jan 28;374(4):384-386 [FREE Full text] [doi: 10.1056/nejme1515172]

XSL•FO
RenderX

10. Data Sharing Policy. Barts and the London School of Medicine and Dentistry. Pragmatic Clinical Trials Unit. URL: https://tinyurl.com/4e79w3j4 [accessed 2022-01-05]
11. Network data sharing policy. NIH StrokeNet. URL: https://www.nihstrokenet.org/docs/default-source/default-document-library/adm04-strokenet-network-data-sharing-15-sep-2016f84cedd0c2c66f0f86baff000016a715.pdf?sfvrsn=0 [accessed 2022-01-05]
12. Bull S, Cheah PY, Denny S, Jao I, Marsh V, Merson L, et al. Best practices for ethical sharing of individual-level health research data from low- and middle-income settings. J Empir Res Hum Res Ethics 2015 Jul;10(3):302-313 [FREE Full text] [doi: 10.1177/1556264615594606] [Medline: 26297751]
13. Baber N. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use (ICH). Br J Clin Pharmacol 1994 May;37(5):401-404 [FREE Full text] [doi: 10.1111/j.1365-2125.1994.tb05705.x] [Medline: 8054244]
14. Data sharing for grants—final research data SOP. National Institute of Allergy and Infectious Diseases. URL: https://www.niaid.nih.gov/research/grants-data-sharing-final-research [accessed 2022-01-05]
15. Pisani E, Aaby P, Breugelmans JG, Carr D, Groves T, Helinski M, et al. Beyond open data: realising the health benefits of sharing data. BMJ 2016 Oct 10;355:i5295 [FREE Full text] [doi: 10.1136/bmj.i5295] [Medline: 27758792]
16. Data management and sharing policy. National Institute of Health Scientific Data Sharing. URL: https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm [accessed 2022-01-05]
17. Blasimme A, Fadda M, Schneider M, Vayena E. Data sharing for precision medicine: policy lessons and future directions. Health Aff (Millwood) 2018 May;37(5):702-709. [doi: 10.1377/hlthaff.2017.1558] [Medline: 29733719]
18. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: practices and perceptions. PLoS One 2011;6(6):e21101 [FREE Full text] [doi: 10.1371/journal.pone.0021101] [Medline: 21738610]
19. Glickman SW, McHutchison JG, Peterson ED, Cairns CB, Harrington RA, Califf RM, et al. Ethical and scientific implications of the globalization of clinical research. N Engl J Med 2009 Feb 19;360(8):816-823. [doi: 10.1056/NEJMsb0803929] [Medline: 19228627]
20. Extension of the Transition Period under Article 66.1 of the TRIPS Agreement for Least-Developed Country Members for Certain Obligations with Respect to Pharmaceutical Products. Council for TRIPS. URL: https://www.wto.org/english/tratop_e/trips_e/art66_1_e.htm [accessed 2022-01-05]
21. Data exclusivity and other “trips-plus” measures. World Health Organization, Office for South-East Asia. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/272979/Data-exclusivity.pdf?sequence=1&isAllowed=y [accessed 2022-01-05]
22. Trips agreement- article 66 (practice). World Trade Organization Analytical Index. URL: https://www.wto.org/english/res_e/publications_e/ai17_e/trips_art66_oth.pdf [accessed 2022-01-05]
23. COVID-19 IP Policy Tracker. World Intellectual Property Organization (WIPO). URL: https://www.wipo.int/covid19-policy-tracker/#/covid19-policy-tracker/ipo-operations [accessed 2022-01-05]
24. Attaran A, Barry D, Basheer S, Bate R, Benton D, Chauvin J, et al. How to achieve international action on falsified and substandard medicines. BMJ 2012 Nov 13;345:e7381. [doi: 10.1136/bmj.e7381] [Medline: 23149211]
25. British Medical Journal moves to flush out secret trial data. Reuters. 2012 Oct 21. URL: https://www.reuters.com/article/pharmaceuticals-disclosure-bmj-idUSL5E8LV59720121031 [accessed 2022-03-15]
26. Sampath PG. India's product patent protection regime: less or more of "pills for the poor"? J World Intellectual Property 2006 Nov;9(6):694-726 [FREE Full text] [doi: 10.1111/j.1422-2213.2006.00308.x]
27. Srinivasan S. The draft national pharmaceuticals policy: concerns relating to data exclusivity and price control. Indian J Med Ethics 2006 Oct 1(4) [FREE Full text] [doi: 10.20529/ijme.2006.047]
28. 't Hoen EF. The global politics of pharmaceutical monopoly power: drug patents, access, innovation. The application of the WTO Doha declaration on TRIPS public health. Amsterdam, The Netherlands: AMB Publishers; Jan 2009.
29. Parker M, Bull S. Sharing public health research data: toward the development of ethical data-sharing practice in low- and middle-income settings. J Empir Res Hum Res Ethics 2015 Jul;10(3):217-224 [FREE Full text] [doi: 10.1177/1556264615593494] [Medline: 26297744]
30. Zink A, Silman AJ. Ethical and legal constraints on data sharing between countries in multinational epidemiological studies in Europe report from a joint workshop of the European League Against Rheumatism standing committee on epidemiology with the "AutoCure" project. Ann Rheum Dis 2008 Jul;67(7):1041-1043. [doi: 10.1136/ard.2007.080598] [Medline: 18180280]
31. Califf RM. Clinical trials bureaucracy: unintended consequences of well-intentioned policy. Clin Trials 2006;3(6):496-502. [doi: 10.1177/1740774506073173] [Medline: 17170032]
32. Institute of Medicine (US). Sharing clinical research data: workshop summary. Washington, DC, USA: National Academies Press (US); Mar 29, 2013.
33. Communication and dissemination of results. MRC Clinical Trials Unit. URL: https://cdsatoolkit.thsti.in/communication-and-dissemination-of-results/ [accessed 2022-01-08]
34. Peterson J, Pearce PF, Ferguson LA, Langford CA. Understanding scoping reviews: definition, purpose, and process. J Am Assoc Nurse Pract 2017 Jan;29(1):12-16. [doi: 10.1002/2327-6924.12380] [Medline: 27245885]

XSL•FO
RenderX

35. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
36. 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] [Medline: 20854677]
37. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Sep 4:467-473 [FREE Full text] [doi: 10.7326/m18-0850]
38. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. BMC Med Res Methodol 2018 Jan 10;18(1):5 [FREE Full text] [doi: 10.1186/s12874-017-0468-4] [Medline: 29316881]
39. COVID-19 Clinical Research Coalition. URL: https://covid19crc.org/ [accessed 2022-01-08]
40. Policies and procedures to guide external investigator access to clinical trial data. the YODA Project. URL: https://yoda.yale.edu/policies-procedures-guide-external-investigator-access-clinical-trial-data [accessed 2022-01-06]
41. Clinical Trials Data Sharing. Celgene. URL: https://www.celgene.com/research-development/clinical-trials/clinical -trials-data-sharing/Accessed [accessed 2022-01-05]
42. Final NIH policy for data management and sharing. National Institute of Health Office of Extramural Research. URL: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html [accessed 2022-01-05]
43. NHLBI Policy for data sharing from clinical trials and epidemiological studies. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/grants-and-training/policies-and-guidelines/nhlbi-policy-for-data-sharing-from-clinical-trials-and-epidemiological-studies [accessed 2022-01-05]
44. NCI Cancer Moonshot public access and data sharing policy. National Care Institute. URL: https://www.cancer.gov/research/key-initiatives/moonshot-cancer-initiative/funding/public-access-policy [accessed 2022-01-05]
45. Facilitating European Clinical Research. European Clinical Research Infrastructure Network. URL: https://ecrin.org/ [accessed 2021-08-18]
46. Access and manage your VA benefits and health care. U.S. Department of Veterans Affairs. URL: https://www.va.gov// [accessed 2021-08-18]
47. SRDR+: moving systematic reviews forward. Systematic Review Data Repository. URL: https://srdrplus.ahrq.gov/?no_redirect=true [accessed 2021-08-18]
48. Heart Attack and Stroke Symptoms. American Heart Association. URL: https://www.heart.org/ [accessed 2021-08-18]
49. Joint statement on public disclosure of clinical trials. World Health Organization. 2017 May 18. URL: https://www.who.int/news/item/18-05-2017-joint-statement-on-registration [accessed 2021-08-18]
50. Medicines and Health care products Regulatory Agency. United Kingdom Government. URL: https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency [accessed 2021-08-18]
51. ClinicalStudyDataRequest.com (CSDR). URL: https://www.clinicalstudydatarequest.com/ [accessed 2021-08-18]
52. World-Wide Antimalarial Resistance Network. URL: https://www.wwarn.org/ [accessed 2021-08-18]
53. U.S. Food & Drug Administration. URL: https://www.fda.gov/ [accessed 2021-08-18]
54. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data sharing statements for clinical trials: a requirement of the international committee of medical journal editors. Ann Intern Med 2017 Jul 04;167(1):63-65 [FREE Full text] [doi: 10.7326/M17-1028] [Medline: 28586790]
55. The final version of the RDA COVID-19 recommendations and guidelines for data sharing. Research Data Alliance. 2020 Jun 30. URL: https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-1 [accessed 2021-08-18]
56. Our Data-Sharing Philosophy. Project Data Sphere. URL: https://www.projectdatasphere.org/about/data-sharing-philosophy [accessed 2021-08-18]
57. Volunteer for COVID-19 Clinical Studies. National Institute of Allergy and Infectious Diseases. URL: https://www.niaid.nih.gov/ [accessed 2021-08-18]
58. External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use. European Medicines Agency. 2017 Sep 20. URL: https://tinyurl.com/yckkukvd [accessed 2021-08-18]
59. PhRMA Principles for Responsible Clinical Trial Data Sharing – Certifications. Pharmaceutical Research and Manufacturers of America. URL: https://www.phrma.org/en/Codes-and-guidelines/PhRMA-Principles-for-Responsible-Clinical-Trial-Data-Sharing-Certifications [accessed 2021-08-18]
60. Open research data: clinical trials and public health interventions. UK Research and Innovation. URL: https://mrc.ukri.org/research/policies-and-guidance-for-researchers/open-research-data-clinical-trials-and-public-health-interventions/ [accessed 2021-08-18]
61. Document library. National Institute for Health Research. URL: https://www.nihr.ac.uk/about-us/document-library.htm?custom_in_type=9697 [accessed 2021-08-18]
62. Pragmatic Clinical Trials Unit. Queen Mary University of London. URL: https://www.qmul.ac.uk/pctu/ [accessed 2021-08-18]
63. Policy for Data Management and Data Sharing. Patient-Centered Outcomes Research Institute. URL: https://www.pcori.org/search/site?keyword=data%20sharing%20policy [accessed 2021-08-18]

XSL•FO
RenderX

64. United Kingdom-Clinical Research Collaboration. URL: https://cdn.ymaws.com/www.ukcrc-ctu.org.uk/resource/resmgr/5_year_strategy/network_5_year_strategy_2019.pdf [accessed 2021-08-18]
65. Mateen FJ, Oh J, Tergas AI, Bhayani NH, Kamdar BB. Titles versus titles and abstracts for initial screening of articles for systematic reviews. Clin Epidemiol 2013;5:89-95 [FREE Full text] [doi: 10.2147/CLEP.S43118] [Medline: 23526335]
66. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
67. Rao TS, Radhakrishnan R, Andrade C. Standard operating procedures for clinical practice. Indian J Psychiatry 2011 Jan;53(1):1-3 [FREE Full text] [doi: 10.4103/0019-5545.75542] [Medline: 21430999]
68. Daudt H, van Mossel C, Scott S. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
69. Sharing individual participant data from clinical trials. UK Research andInnovation. 2015 Apr 23. URL: https://www.ukri.org/publications/sharing-individual-participant-data-from-clinical-trials/ [accessed 2022-01-08]
70. Ross JS, Waldstreicher J, Bamford S, Berlin JA, Childers K, Desai NR, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. Sci Data 2018 Nov 27;5:180268 [FREE Full text] [doi: 10.1038/sdata.2018.268] [Medline: 30480665]
71. Mitka M. Clinical trial data: share and share alike? JAMA 2015 Mar 03;313(9):881-882. [doi: 10.1001/jama.2015.481] [Medline: 25734721]
72. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. N Engl J Med 2013 Aug 01;369(5):475-478. [doi: 10.1056/NEJMsr1302541] [Medline: 23902490]
73. Pencina MJ, Louzao DM, McCourt BJ, Adams MR, Tayyabkhan RH, Ronco P, et al. Supporting open access to clinical trial data for researchers: the Duke Clinical Research Institute-Bristol-Myers Squibb supporting open access to researchers initiative. Am Heart J 2016 Feb;172:64-69. [doi: 10.1016/j.ahj.2015.11.002] [Medline: 26856217]
74. Principles for responsible clinical trial data sharing – certifications. Pharmaceutical and Research Manufacturers of America. URL: https://phrma.org/resource-center/Topics/Clinical-Trials/PhRMA-Principles-for-Responsible-Clinical-Trial-Data-Sharing-Certifications [accessed 2022-01-05]
75. NIHR position on the sharing of research data. National Institute of Health and Care Research. 2019 May 20. URL: https://www.nihr.ac.uk/documents/nihr-position-on-the-sharing-of-research-data/12253 [accessed 2022-01-05]
76. MRC policy and guidance on sharing of research data from population and patient studies. Medical Research Council. URL: https://www.ukri.org/wp-content/uploads/2021/08/MRC-0208212-MRC-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies-Word-version-v01.02.pdf [accessed 2022-01-05]
77. Policy for data management and data sharing. Patient-Centered Outcome Research Institute. URL: https://www.pcori.org/sites/default/files/PCORI-Policy-for-Data-Management-and-Data-Sharing.pdf [accessed 2022-01-05]
78. Considerations for a participant data sharing SOP. UKCRC Registered Clinical Trials Units. URL: https://ukcrc-ctu.org.uk/wp-content/uploads/2021/04/data_sharing_sop_guide_v1.1_-1.pdf [accessed 2022-01-05]
79. European medicines agency policy on publication of clinical data for medicinal products for human use. European Medicines Agency. 2019 Mar 21. URL: https://www.ema.europa.eu/en/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf [accessed 2022-01-05]
80. Chakrabarti G. Need of data exclusivity: impact on access to medicine. J Intellect Prop Rights 2014 Sep;19(5):325-336.
81. Haug CJ. From patient to patient--sharing the data from clinical trials. N Engl J Med 2016 Jun 23;374(25):2409-2411. [doi: 10.1056/NEJMp1605378] [Medline: 27168009]
82. Rockhold FW. Statistical controversies in clinical research: data access and sharing-can we be more transparent about clinical research? Let's do what's right for patients. Ann Oncol 2017 Aug 01;28(8):1734-1737 [FREE Full text] [doi: 10.1093/annonc/mdx123] [Medline: 28383637]
83. Lo B. Sharing clinical trial data: maximizing benefits, minimizing risk. JAMA 2015 Feb 24;313(8):793-794. [doi: 10.1001/jama.2015.292] [Medline: 25594500]
84. Mbuagbaw L, Foster G, Cheng J, Thabane L. Challenges to complete and useful data sharing. Trials 2017 Feb 14;18(1):71 [FREE Full text] [doi: 10.1186/s13063-017-1816-8] [Medline: 28196540]
85. Rowhani-Farid A, Allen M, Barnett AG. What incentives increase data sharing in health and medical research? A systematic review. Res Integr Peer Rev 2017;2:4 [FREE Full text] [doi: 10.1186/s41073-017-0028-9] [Medline: 29451561]

Abbreviations

GSK: GlaxoSmithKline
IPD: individual participant data
IPR: intellectual property rights
LMIC: low- and middle-income country
NIH: National Institutes of Health

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
WTO: World Trade Organization

Edited by A Mavragani; submitted 15.09.21; peer-reviewed by M Vivalya, L Suppan; comments to author 15.10.21; revised version received 09.01.22; accepted 21.03.22; published 04.05.22.

Please cite as:
Gudi N, Kamath P, Chakraborty T, Jacob AG, Parsekar SS, Sarbadhikari SN, John O
Regulatory Frameworks for Clinical Trial Data Sharing: Scoping Review
J Med Internet Res 2022;24(5):e33591
URL: https://www.jmir.org/2022/5/e33591
doi:10.2196/33591
PMID:35507397

©Nachiket Gudi, Prashanthi Kamath, Trishnika Chakraborty, Anil G Jacob, Shradha S Parsekar, Suptendra Nath Sarbadhikari, Oommen John. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2022. 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.

XSL•FO
RenderX

Review

Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review

Timothy Charles Kariotis[1,2], BSocSc, BNutrSc, BA, MPH; Megan Prictor[3,4], LLB, PhD; Shanton Chang[1], BCom, PhD; Kathleen Gray[4], MLibSc, MEnvSc, PhD

[1]School of Computing and Information Systems, University of Melbourne, Parkville, Australia

[2]Melbourne School of Government, The University of Melbourne, Carlton, Australia

[3]Melbourne Law School, University of Melbourne, Carlton, Australia

[4]Centre for Digital Transformation of Health, University of Melbourne, Parkville, Australia

Corresponding Author:
Timothy Charles Kariotis, BSocSc, BNutrSc, BA, MPH
School of Computing and Information Systems
University of Melbourne
The University of Melbourne
Parkville, 3052
Australia
Phone: 61 0488300223
Email: Timothy.kariotis@unimelb.edu.au

Related Article:

Comment in: http://mhealth.jmir.org/2022/5/e38513/

Abstract

Background: The adoption of electronic health records (EHRs) and electronic medical records (EMRs) has been slow in the mental health context, partly because of concerns regarding the collection of sensitive information, the standardization of mental health data, and the risk of negatively affecting therapeutic relationships. However, EHRs and EMRs are increasingly viewed as critical to improving information practices such as the documentation, use, and sharing of information and, more broadly, the quality of care provided.

Objective: This paper aims to undertake a scoping review to explore the impact of EHRs on information practices in mental health contexts and also explore how sensitive information, data standardization, and therapeutic relationships are managed when using EHRs in mental health contexts.

Methods: We considered a scoping review to be the most appropriate method for this review because of the relatively recent uptake of EHRs in mental health contexts. A comprehensive search of electronic databases was conducted with no date restrictions for articles that described the use of EHRs, EMRs, or associated systems in the mental health context. One of the authors reviewed all full texts, with 2 other authors each screening half of the full-text articles. The fourth author mediated the disagreements. Data regarding study characteristics were charted. A narrative and thematic synthesis approach was taken to analyze the included studies' results and address the research questions.

Results: The final review included 40 articles. The included studies were highly heterogeneous with a variety of study designs, objectives, and settings. Several themes and subthemes were identified that explored the impact of EHRs on information practices in the mental health context. EHRs improved the amount of information documented compared with paper. However, mental health–related information was regularly missing from EHRs, especially sensitive information. EHRs introduced more standardized and formalized documentation practices that raised issues because of the focus on narrative information in the mental health context. EHRs were found to disrupt information workflows in the mental health context, especially when they did not include appropriate templates or care plans. Usability issues also contributed to workflow concerns. Managing the documentation of sensitive information in EHRs was problematic; clinicians sometimes watered down sensitive information or chose to keep it in separate records. Concerningly, the included studies rarely involved service user perspectives. Furthermore, many studies provided limited information on the functionality or technical specifications of the EHR being used.

XSL•FO
RenderX

Conclusions: We identified several areas in which work is needed to ensure that EHRs benefit clinicians and service users in the mental health context. As EHRs are increasingly considered critical for modern health systems, health care decision-makers should consider how EHRs can better reflect the complexity and sensitivity of information practices and workflows in the mental health context.

(J Med Internet Res 2022;24(5):e30405) doi:10.2196/30405

KEYWORDS

electronic health records; psychiatry; mental health; electronic medical records; health informatics; mental illness; scoping review; clinical decision support

Introduction

Background

Electronic health records (EHRs) are being adopted in many health systems to improve the collection, sharing, and use of health care information [1]. Such information practices play a critical role in providing safe and high-quality care [2,3]. EHRs promise more integrated and connected health services, which are recognized by the World Health Organization and many governments as essential for sustainable, effective health systems [4-6]. Owing to the complex array of services that support service users, the fragmentation of care and limited information sharing are common in the mental health context [7]. Limited information sharing among health care services affects the planning and provisioning of appropriate care, such as medication management and reconciliation [8,9]. It can also negatively affect service users' experience of mental health care, especially when it leads to them having to retell their stories multiple times [10]. However, information sharing also comes with risks for service users, such as the stigma associated with mental health conditions [11]. Thus, mental health information tends to be considered highly sensitive information, requiring extra protection [12].

Information is critical to modern health care, especially mental health care, and health records are vital tools for documenting, organizing, and using information [8,13]. When health care professionals provide care to service users, they undertake a range of information practices, including seeking, using, documenting, and sharing information [14]. Health records play a critical role in such practices. Coiera [15] outlined that a health record has many functions, including enabling communication among staff through the information in the record, providing a central source of information for care, acting as an informal workspace for capturing ideas, and being a historical archive that can inform future care. Mental health records are especially complex because many entries can be included in the record [16,17].

EHRs are a core health informatics tool for the improvement of health care quality, partly through improved information quality and accessibility [15]. EHRs are, in one sense, a digitized version of the health care record but are also much more in that they introduce new practices and workflows [18-21]. For example, EHRs have been found to affect how information is documented in clinical records by introducing structured data entry forms and disrupting the collection of narrative information [22-25]. Internationally, the uptake of EHRs in the mental health context has been much slower than in other health contexts [26-29]. A recent scoping review on the effective implementation of electronic medical records (EMRs) in mental health settings also identified limited research on this topic [30]. Apart from the barriers faced by all health settings in adopting EHRs, such as interoperability, time impacts, and workflow changes, there may be particular issues in the mental health context that require investigation [31].

Information sharing relies on a range of information behaviors and practices by clinicians and service users [32,33]. Information behavior has been used to capture the range of human behaviors related to seeking and using information [34]. In comparison, information practice considers how information behaviors are embedded and shaped by organizational contexts and interactions [34]. Østensen et al [35] defined information practice as "a socially constructed practice that determines how information is produced, organised, disseminated, distributed, reproduced and circulated in the community, and which specific types of information are legitimized."

Going forward, we purposively use the term *information practice* rather than the more widely used term *information behavior*. Adopting this language aligns with our understanding that social and organizational rules and norms shape how clinicians practice information sharing [36-40]. Using the concept of information practice allows us to reflect on how particular issues in the mental health context, such as sensitive information and stigma, influence information practices.

Mental health care involves various sensitive information practices, such as people sharing a range of sensitive and potentially stigmatizing information, from personal trauma to behavioral patterns [9,41]. This information can also be considered stigmatizing, both publicly and within health care settings. Stigma is a common theme across a number of studies exploring the experiences of service users in the mental health context [42-44]. For example, it has been found that people with diagnoses such as borderline personality disorder experience stigma from health professionals, which affects their care [45]. Health care professionals in the mental health context are also aware of the sensitivity of mental health information [9,46]. Several commentaries have raised concerns about how sensitive information is recorded in EHRs and its implications for privacy and security [47-50].

The documentation of mental health information is another information practice that is an issue in EHR use in the mental health context. Mental health services are more likely to rely on narrative information [51]. For example, Kobus et al [51] pointed out that although most medical conditions rely on

XSL•FO
RenderX

quantitative measures, depression relies partly on reviewing narrative progress notes. However, one of the reasons for adopting an EHR is to standardize data collection through structured data fields [24]. The lack of standardized information formats in the mental health context is a potential barrier to EHR uptake [52,53]. There is also great diversity in how mental health information is documented and used across professions, which complicates the standardization of mental health information [54]. Although diagnostic codes are available for mental health conditions, it is not easy to establish a clear diagnosis and associated diagnostic codes in the mental health context [55-57].

A final issue that has been raised in the literature as a concern for the adoption of EHRs in mental health contexts is the impact it could have on the therapeutic relationship [58,59]. Therapeutic relationships are critical for providing mental health care [60,61]. Adding an EHR to clinical encounters, which may bring new information practices, has been raised in different care settings as a potential barrier to establishing and maintaining a therapeutic relationship [62]. Shank et al [63] found that mental health clinicians worried that EHRs would divert their attention from service users and negatively affect the therapeutic relationship.

Research on the use of EHRs in the mental health context is at a low stage of maturity, with a diverse array of studies responding to different contextual issues. Thus, a scoping review is necessary to understand the literature [64]. This scoping review aims to identify the impact of EHRs, implemented in the mental health context, on information practices. Furthermore, it aims to explore how, in the use of EHRs, sensitive information, data standardization, and impacts on the therapeutic relationship have been considered, if at all.

The review had the following objectives and research questions:

1. In mental health contexts, what impact do EHRs have on information practices, and how do these changes affect other aspects of care?
2. In mental health contexts, how have sensitive information, data standardization, and therapeutic relationships been managed when using EHRs?

A Note on Language

We chose to use the term *service user* to represent people accessing and using mental health services and chose not to use terms such as *client* as this suggests that people voluntarily use services, which is not always the case in mental health contexts. Terms such as *patient* can be considered as disempowering for people who access services. We acknowledge that the terminology in this space is not settled and that others may consider different terms more appropriate.

The title of this paper refers to the *mental health context*. We chose this term to capture the broad range of clinical and nonclinical services that people may access when experiencing mental health issues [65].

Throughout this paper, we have raised terms such as *mental health data* and *mental health information*. These terms are not clearly defined in the literature, and we will return to this issue in the *Discussion* section.

Methods

Overview

The scoping review is a method of synthesizing research and can support various methods, objectives, and study types [64,66,67]. Unlike systematic reviews, scoping reviews do not attempt an exhaustive review of all relevant studies but rather aim for a breadth of evidence. Owing to the relatively recent uptake of EHRs in mental health care, it is appropriate to conduct a scoping review of this emerging evidence to consider a broad definition of EHRs and a range of study types.

This scoping review was informed by the Arksey and O'Malley [67] framework for scoping reviews. We were also informed by the PRISMA-ScR (Preferred Reporting Item for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist and explanation [68]. However, some criteria were not relevant to our study because of the thematic synthesis approach we used to analyze the included studies. Our approach to this scoping review has been to explore the literature on EHRs and describe what it tells us about the impact of EHRs on information practices in the mental health context. Unlike some scoping reviews, we chose not to map the trends in the literature. As different jurisdictions are moving at different speeds in their adoption of EHRs, and due to the breadth of the topic, we did not view the mapping of trends as feasible or helpful in this specific review.

Inclusion and Exclusion Criteria

Overview

We included studies that have examined EHRs in the adult mental health context, either by being based in mental health settings or being used by or for people with a mental health diagnosis. Nonclinical services (eg, housing services) providing services to people with mental illness were also included in this review, in keeping with the definition of health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" [69].

We included studies that mentioned using EMRs, EHRs, or any associated terms such as health information systems. Häyrinen et al [70], in a review of the literature, found that there are many terms used to describe EHRs, with various functions, formats, users, settings, and purposes. We acknowledge that EHRs, EMRs, and other terms are different but interlinked systems. An EMR is generally considered to be a record of a person's health encounters in a specific health setting. In contrast, an EHR is usually a compilation of summary information from across EMRs in a region, country, or health system [71]. However, these definitions are not always made clear or defined in the literature, and thus, we did not adopt a specific definition in this paper. There is no one gold standard definition of an EHR or EMR, with peak health informatics organizations using the same definition for both terms [72]. Going forward, we have used the term EHR as an umbrella term to represent the

XSL•FO
RenderX

information systems used to manage service users' health information by and for health services.

This review included any primary evidence that explored the use of EHRs in the mental health context published before April 2021. We excluded studies that focused on children's health care in acknowledgment that this field raises several unique issues, such as the involvement of parents, which is worthy of a specific review. We excluded studies during the full-text screening that were not relevant to EHRs, the mental health context, or information practices. Studies that focused on clinicians' perceptions of EHRs in general rather than the EHR that was implemented were also excluded. The case studies were evaluated on a case-by-case basis, depending on the level of detail provided. We excluded studies that described only the design and development of an EHR.

Several types of EHRs provide service users access to their health information, such as personal EHRs, patient portals, and initiatives such as OpenNotes. We excluded these from this review as they raise unique issues regarding how service users access and use their health information. We acknowledge that systems such as OpenNotes will have implications for our study questions. However, we consider these systems to be more thematically aligned with patient portals and personal EHRs, which would benefit from a separate review. Readers interested in this topic should read the recent scoping review by Zhang et al [73] on the use of patient portals in mental health settings.

Types of Studies, Information Sources, and Search Strategy

Embase, Scopus, and PsycINFO were searched using a combination of key terms, an example of which is provided in Textbox 1. The search strategy was developed iteratively alongside the identification of key terms in the literature and hand searching of reference lists. This search was initially undertaken in late 2018 and then updated in December 2020, with new papers continually identified until April 2021, when the final draft was completed. No date limitation was applied in the initial search as we wanted to identify all relevant health informatics literature, which ranged across several decades [74]. Papers not published in English were excluded. The first author (TCK) read a subset of articles from the initial search to develop further search terms, which were then applied across PubMed, CINAHL, SocINDEX, and Web of Science. We also searched research repositories: Google, Google Scholar, Grey Literature Report, TROVE, OPEN Grey, and Social Care Online.

Textbox 1. Example search strategy run on Embase.

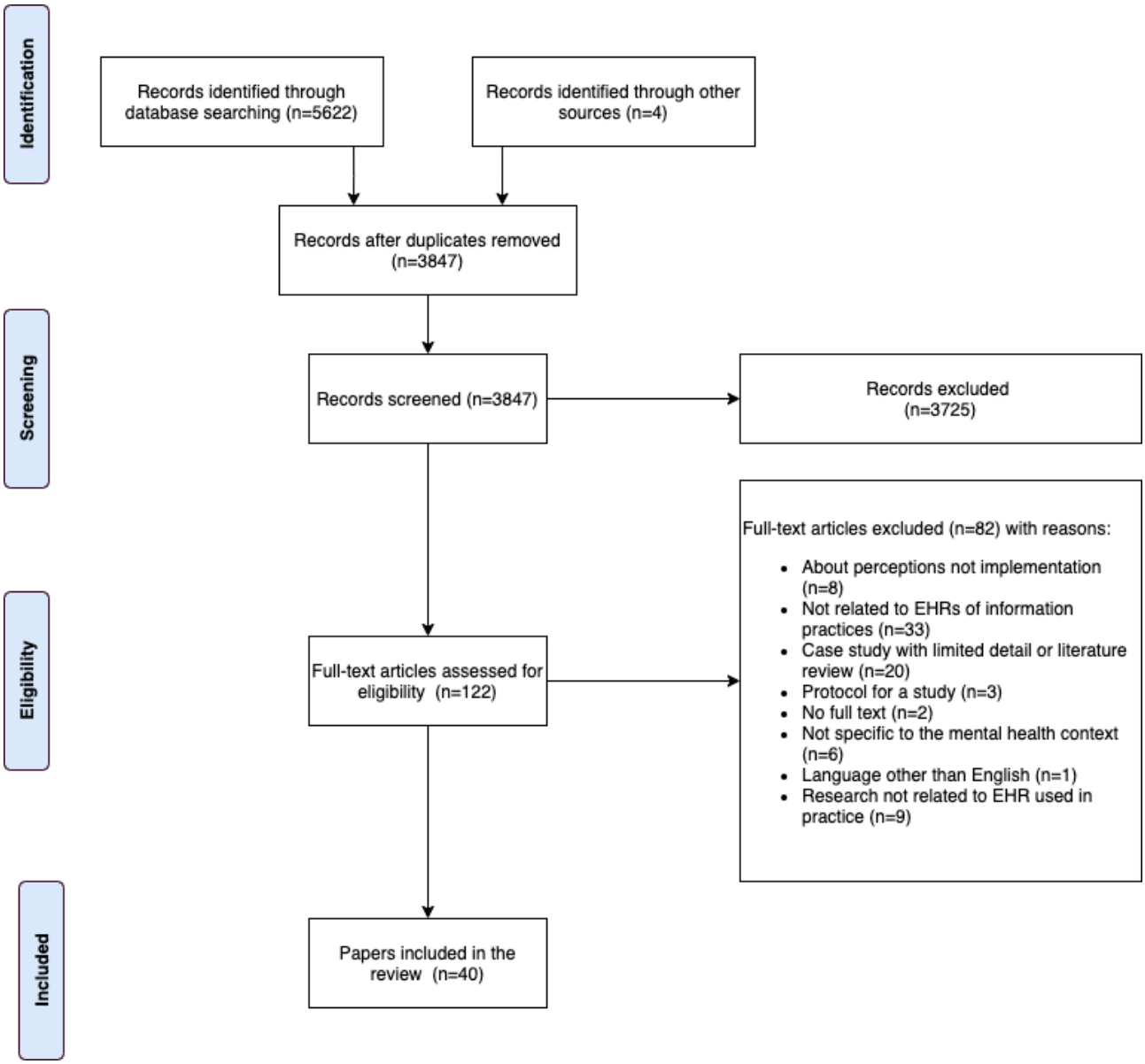
Search number and search term
1. Electronic health record/
2. Electronic medical record*
3. Electronic patient record*
4. EHR
5. EPR
6. Health information system
7. Health Information Exchange
8. Mental Disease
9. Mental Illness
10. Mental health care
11. Behavio?l health care
12. Mental health service*
13. 1 OR 2 OR 3 OR 4 OR 5 OR 6
14. 7 OR 8 OR 9 OR 10 OR 11
15. 12 AND 13

Study Selection

We identified 3847 nonduplicate articles. The titles and abstracts were screened against the inclusion criteria by TCK. Approximately 3.17% (122/3847) of articles were considered potentially relevant and were retrieved from the full text. TCK reviewed all 122 full-text articles, with SG and MP each independently reviewing half of the full-text articles. Differences were resolved by discussion and mediation by KG. Of the 122 articles reviewed in the full text, 82 (67.2%) were excluded, and 40 (32.8%) were included (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) diagram. EHR: electronic health record.



Charting the Results

To provide an overview of the study characteristics, we charted the objectives, study design, research method, study participants, country, study setting, EHR description, year of publication, and theories used in the included studies in a spreadsheet.

Owing to the breadth of the study types and objectives, covering a range of qualitative and quantitative methodologies, we followed Lakshman et al [75] in adopting both a textual narrative and thematic synthesis approach to analyzing the included studies. The textual analysis involved tabulating the study findings alongside their characteristics and conclusions. We adopted a thematic synthesis approach to analyze the qualitative papers included in this review. This method involved coding the text, developing descriptive themes, and generating analytical themes [76,77]. Following the method by Thomas and Harden [76], we initially developed descriptive themes by coding both direct participant quotes and researcher interpretations. We approached this by free-coding findings in an unstructured mind map, which we used to develop descriptive themes. Our research questions framed this coding process so that we coded anything related to information practices or the issues outlined in the *Introduction* section. We also considered the factors that shaped the way information was collected or shared in the presence of an EHR or the outcomes of changes in information practices. We intended for these descriptive themes to stay as close to the original findings as possible.

Next, we used our review questions to develop the analytical themes. We integrated the quantitative data we had extracted during the textual analysis during this process. Thomas and Harden [76] described this process as potentially controversial as it relies on the researcher's judgment and insight. This iterative process aimed to capture the descriptive data developed in the initial analysis. The analysis identified 6 major themes describing the impact of EHRs on information practices in the mental health context. Of these themes, 5 had subthemes that explored specific topics relevant to the theme.

XSL•FO
RenderX

Study Characteristics

Overview

The studies included in this review were highly heterogeneous. This heterogeneity reflects one of the strengths of a scoping review in that it was inclusive of many study types. The following sections describe the characteristics of the included studies. Owing to the heterogeneity of study types and limited use of standardized terms, comparisons between studies were limited. Table 1 outlines the key study characteristics.

Table 1. Study characteristics (N=40).

Characteristics	Values, n (%)
Study design	
Quantitative	21 (53)
Qualitative	11 (28)
Mixed methods	8 (20)
Research method[a]	
Surveys	15 (38)
Interviews or focus groups	7 (18)
Chart reviews	5 (13)
Cross-sectional or secondary data use	5 (13)
Quality improvement	3 (8)
Ethnographic or observational	6 (15)
Descriptive case studies	1 (3)
Study sample	
Clinicians or health care professionals	26 (65)
Administrator, IT[b], or management	9 (23)
Service users	4 (10)
No participants (eg, record review)	13 (33)
Countries	
United States	27 (68)
United Kingdom	8 (20)
Canada	2 (5)
Other	3 (8)

[a]Some studies included multiple methods and thus were counted twice.

[b]IT: information technology.

Study Design and Research Method

A range of study designs and research methods were represented in the included studies. Most were quantitative (21/40, 53%) [78-98], with qualitative (11/40, 28%) [99-109] and mixed method studies (8/40, 20%) [110-117] also included. We categorized studies based on the broad category of research methods, including surveys (15/40, 38%) [78,80-85,89,92,95-97,110,113,116], qualitative interview/focus group studies (7/40, 18%) [100-104,110,111], chart review of specific EHRs (5/40, 13%) [85-87,114,117], cross-sectional analysis of EHR data or comparison with other secondary data (5/40, 13%) [90,91,93,95,98], quality improvement initiatives (3/40, 8%) [79,88,111], ethnographic or observational (6/40, 15%) [99,105-108,112], and descriptive case studies (1/40, 3%) [109].

The objectives of the included studies varied. We compared the objectives and research questions of the included studies and grouped them according to similar topic areas, as outlined in Table 2 (some studies had multiple objectives). We also included the publication years in Table 2 to showcase how certain topics were not confined to any specific period.

XSL•FO
RenderX

Table 2. Topics of included studies and related publication dates.

Topics of included studies	Publication years of included studies
Exploring the adoption of EHRs[a] in the mental health care context	2015 [116] and 2018 [94]
Evaluation of an EHR implementation	2009 [78], 2010 [107], 2011 [108], 2012 [99], 2017 [79], and 2018 [110]
Exploring the use of EHRs to provide mutual access to psychiatric records	2013 [80] and 2015 [81]
Exploring the impact of EHRs on the therapeutic relationship or person-centered care	2010 [82], 2011 [84] 2017 [101], 2019 [111], 2020 [83], and 2020 [85]
Exploring the use of EHRs in integrated or collaborative care contexts	2012 [113], 2012 [113], 2015 [112], 2015 [81], and 2018 [86]
Comparing documentation in EHRs with documentation in paper records	2007 [87], 2016 [88], and 2018 [114]
Exploring service users' experiences or satisfaction with care when an EHR is present	2018 [110] and 2020 [90]
Exploring the barriers, facilitators, workarounds, and usability of EHRs in the mental health context	2010 [103], 2011 [108], 2012 [113], 2012 [99], 2013 [109], 2014 [100], 2015 [116], 2015 [112], 2017 [115], 2017 [101], and 2021 [102]
Exploring the impact of EHRs on health care professionals' information practices and behavior	2004 [105], 2010 [106], and 2016 [104]
Exploring clinicians' satisfaction and perspectives of EHRs	2009 [89], 2015 [92], and 2018 [110]
Exploring information availability or documentation of specific diagnoses in EHRs	2013 [117], 2016 [91], 2016 [95], 2016 [96], 2019 [93] 2020 [98], and 2020 [97]

[a]EHR: electronic health record.

Participants

In most studies that involved direct data collection from human participants, such as EHR evaluations, the participants were health care professionals. The type of health care professional was not always reported or was generalized as *medical professionals*. Overall, primary health care clinicians, physicians, psychiatrists, psychologists, social workers, and nurses were well-represented across the studies. Some studies (9/40, 23%) included administrative, management, or information technology staff [78,97,99-103,108,112]. Only 10% (4/40) of studies involved service users [82,83,108,110]. Table 3 provides more details regarding the types of participants in the included studies.

Table 3. Participant roles reported in included studies (N=40).

Participant role	Included studies reporting this role, n (%)	Reference
Primary health care professional	4 (10)	[81,97,110,112]
Physician	6 (15)	[92,99,102,106,107,115]
Psychiatrist	7 (18)	[80,89,96,99-101,103]
Psychologist or psychology technicians	9 (23)	[78,89,92,96,103-105,111,116]
Behavioral health clinicians or mental health clinicians	5 (13)	[83,101,102,110,112]
Nurse, psychiatric nurse, or nurse practitioner	11 (28)	[78,84,89,96,99,100,102,103,105,114,115]
Social workers or social assistants	7 (18)	[92,96,100,103-105,111]
Pharmacists	3 (8)	[78,99,102]
Other allied health professionals	5 (13)	[99,100,105,107,111]
Other clinical or health care staff	12 (30)	[78,85,89,96,97,99-101,107,108,110,111]
Administrative staff	5 (13)	[78,102,103,107,112]
Information technology staff	4 (10)	[97,99,100,108]
Implementation teams	4 (10)	[99,100,107,108]
Service users	4 (10)	[82,83,108,110]
No participants (eg, secondary data and chart review)	13 (33)	[79,86-88,90,91,93-95,98,109,113,117]

XSL•FO
RenderX

Countries

Most studies took place in the United States (27/40, 68%) [80-83,85-90,92-98,101,102,104,106,110-113,115,116], followed by the United Kingdom (8/40, 20%) [78,84,99,100,105,107,108,117]. Canada (2/40, 5%) [79,114], France (1/40, 3%) [103], Brazil (1/40, 3%), and Ireland (1/40, 3%) [91] were also represented in the included studies. There were no clear differences in the approaches or methods across jurisdictions. The limited number of studies in countries outside the United States and the heterogeneity of study types limited any comparison.

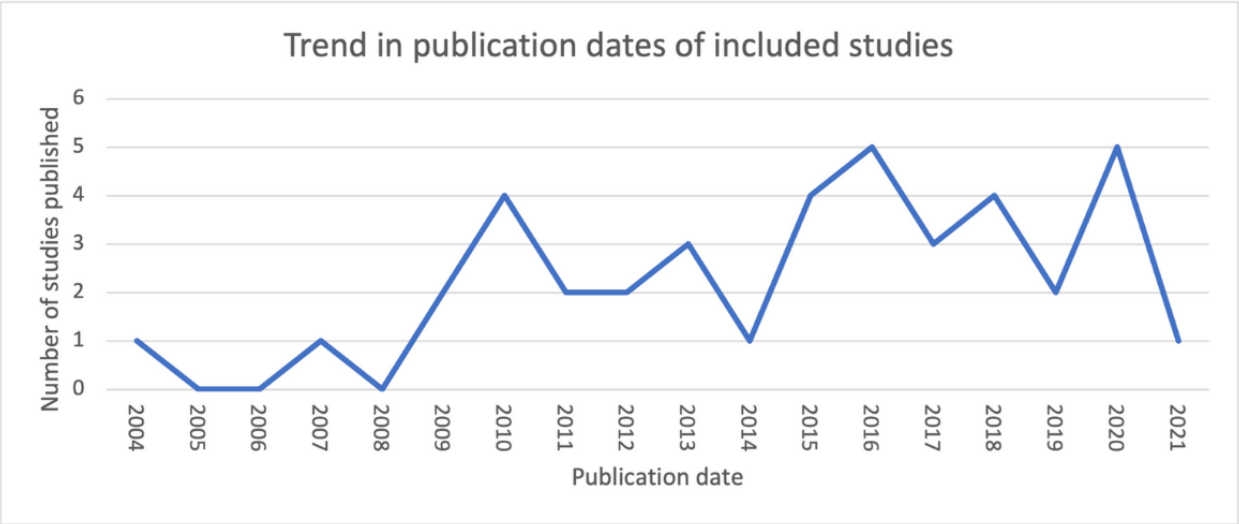
Settings

A variety of health care settings were represented in the included studies, ranging from psychiatric hospitals to community mental health settings. The type of setting was not reported to support easy comparison. These settings are outlined in Multimedia Appendix 1 [54,77-113,115,116] using terminology from the included studies.

Year of Publication

The included studies ranged in publication date from 2004 to 2021 (Figure 2). Although our search strategy had no date restrictions, the terminology used in the search may have shaped what studies were included. Older systems such as computerized patient records may not have been identified. This search strategy was deemed appropriate as these systems did not align with the more recent conceptualizations of what an EHR includes. In general, there has been an increase in the literature on this topic since 2004. Interestingly, many of the issues and topics identified in the *Results* section do not appear to be constrained to a certain period. We would expect to see advancements in EHR infrastructure being reflected in the themes and issues raised in the included studies. This lack of visible change may be because of the low reporting of EHR functions and technical features, limiting the opportunity to see major trends in how EHRs have advanced over time in the mental health context. Table 2 outlines the key topics and publication dates of the included studies.

Figure 2. Trend in publication year of the included studies.



EHR System Used

We noted the name of the EHR and whether it was custom built or off the shelf. We also assessed whether the EHR functions or technical details had been reported. We did not expect all the studies to report this information, such as studies drawing on the secondary analysis of data. We first identified studies that expected to report the details of the EHR in their methods sections, such as the evaluations of specific EHRs (28/40, 70%) [78,79,81-84,87-89,91,92,95,99-111,114,115,117]. Of these 28 studies, 16 (57%) either named the EHR or provided details as to whether it was custom built or off the shelf [78,79,83,84,89,95,99-102,105-108,110,117]. One of the studies pointed to other publications in which the details of the EHR were reported [103]. Of the studies that reported details of the EHRs used, we tried to establish whether they were off-the-shelf commercial EHRs or custom-built EHRs. Some papers did not provide these details, and we searched for further information on the web to categorize those EHRs. Of the 16 studies that reported on the EHR, 3 (19%) were custom built [89,106,117], 12 (75%) were off the shelf [78,79,83,84,95,99-102,107,108,110], and 1 (6%) was unclear [105]. The common off-the-shelf models were RiO [84,99,100,107,108] and EPIC [95,110]. Some studies outlined that commercial off-the-shelf EHRs have been adapted for the mental health context, such as through the addition of mental health–specific modules [102,110]. However, most studies did not clearly state whether and how off-the-shelf EHRs had been customized for the local context.

Of the 28 studies that we expected to report EHR details in their methods sections, only 7 (25%) discussed the functions of the EHR [79,81,84,89,103,105,110]. Sometimes, functions could be assumed from the results sections. No studies reported on

XSL•FO
RenderX

the technical aspects of an EHR. Owing to the limited reporting of EHR types and functions, a comparison across studies was not feasible. The only theme that arose from these studies was that in the United Kingdom, many National Health System services used the same EHR (RiO), whereas, in the United States, there was more variety.

Several studies involved the collection of data on the type of EHRs that services were using, such as cross-sectional surveys of health services. We expected these studies to report details of the EHR in their results sections (8/40, 20%) [80,85,96,97,112,113,116]. Of these 8 studies, 3 (38%) reported on the names or types of EHR used by the included health services [97,112,116], and 2 (25%) reported some details of the functions [80,112], which were mainly related to those who could access the EHR. The included studies reported a variety of off-the-shelf and custom EHRs. For example, in the survey by Cellucci et al [116], they found that most psychology clinics used a commercial system, whereas a small number used custom-designed systems. Another example was the survey by Wu et al [97], which found 17 different off-the-shelf EHRs used by services in the US National Drug Abuse Treatment Clinical Trials Network. Multimedia Appendix 1 lists the EHRs reported in each study.

Theory

The included studies rarely referred to any underlying theory being used. Of the 40 studies, 2 (5%) of studies that reported using theories from the field of information behavior [104,105], 3 (8%) studies reported using sociotechnical theories [99,107,108], and 2 (5%) studies used the Technology Acceptance Model [110,116]. Approximately 3 (8%) of the 40 other studies also discussed the use of different theories [78,79,115].

Quality of the Studies

Scoping reviews do not incorporate an evaluation of the quality of the included studies, although some authors may consider it appropriate to do so [68]. This scoping review included a diversity of studies that no single evaluation method could appropriately address. A significant quality issue that we identified was the lack of detail regarding the EHRs, such as their functionality. The quality criteria for health informatics papers by Talmon et al [118] recommend studies that include information about the system in use.

Results

Overview

In the following sections, we report the findings of the textual narrative and thematic synthesis of the 40 included studies. The analysis led to the development of 6 main themes and several subthemes. The quotes that support the themes are provided in Multimedia Appendix 2 [79,81,86,88,98-100,104,105,107,110,111].

Supports Better Management of Most Information

This theme relates to how EHRs were found to support certain information practices, such as documentation and information accessibility. However, although EHRs show improvements over paper records, there are still issues with the completeness of documentation, especially of mental health information.

Documentation of Information

Several studies reported an improvement in the completeness of documentation in EHRs compared with paper records [78,87-89,102,113-115]. Electronic documentation also addressed issues of legibility that were common in paper records [88,89,99,108]. Improved documentation may partly be because of EHRs promoting accountability in documentation practice and prompting clinicians for certain information [87,105,114]. Approximately 2 (5%) of the 40 studies suggested a greater coupling of policy and practice guidelines within EHRs compared with paper records as the guidelines could be embedded in the EHR, such as through templates [79,105]. These templates provided less discretion regarding how information collection policies were followed. Although EHRs improved documentation compared with paper records, they still showed poor documentation of certain information [87,95]. Tsai and Bond [87] found that past psychotropic medications, prior hospitalizations, and clinical outcomes were regularly missing in EHRs. Bell et al [117], in scanning an EHR to identify drug- and alcohol-related issues, discovered that relevant information was more likely to be found in free-text progress notes, although structured forms were available. An interesting issue raised by participants in the study of EHR use in an integrated care trial by Cifuentes et al [112] was that new types of health care professionals could bring new types of data that the EHR was not designed to collect.

Some studies have found that EHRs create conditions that might negatively affect the documentation. Ser et al [100] found, in interviews with staff across 2 mental health hospitals, that long delays can occur between information collection and documentation in the EHR. Meredith [78] found, from a survey of community mental health teams, that both an EHR and paper record were used side by side, leading to some information not being documented in the EHR.

The benefits of improving documentation came with an increased time burden for clinicians [100,101,103,111,113]. This time burden was related to issues such as simple documentation tasks requiring multiple steps in the EHR [103]. Matthews [101] found that templates may speed up documentation but create challenges if clinicians need to navigate multiple screens and menus. Increased time spent documenting information in EHRs may lead to time savings when reviewing clinical notes in the future [81,110]. For example, Bhe et al [81] reported that 97% (28/29) of primary care physicians who had received access to psychiatric notes in the EMR reported increased efficiency in encounters with psychiatric service users.

Missing Mental Health Data

Several studies found that mental health information was regularly missing from EHRs, documented in the wrong place, or underdocumented in specific contexts [93,95-98,106]. For example, Gleeson et al [91] found that relying on diagnostic codes in an EHR would have missed 92.4% (110/119) of the mental health diagnoses. However, the information needed to

XSL•FO
RenderX

make a diagnosis was available in other parts of the record. The same issue was found in the US Veterans Affairs EHR, where 40.9% (45/110) of people with a posttraumatic stress disorder diagnosis did not have it recorded [96]. Similarly, Madden et al [95] found that many psychiatric services for people with diagnoses of depression or bipolar disorder were missing from the EHR data when compared with health insurance claims. Gibson et al [104], in exploring how clinicians search for information in an EHR, found that when information is not present, clinicians may assume the opposite. For example, if the information on noncompliance with treatment is not present, clinicians may assume that the service user is compliant.

There are many reasons why mental health information may be missing in EHRs. Zhou et al [106] found that psychosocial information may be communicated verbally between clinical team members and not recorded in an EHR immediately, if at all. This practice may be because of psychosocial information being viewed as too subjective to be initially recorded in the EHR [106]. Wu et al [97] found that substance use disorders were not thoroughly captured in EHRs, partly because of the continued use of paper records for that specific part of the health service. Furthermore, in non–mental health services, mental health–related information collection may occur informally and may not be officially recorded in the EHR [106]. Madden et al [95] found that missing mental health data could result from service users seeking mental health care outside their regular health service. Missing information may also be because of the stigma, as discussed further in the following sections.

Access and Availability of Information

The use of EHRs appeared to improve legibility, timely access, and the availability of information [84,87,89,99,108,109,115]. These improvements allowed information to be found more easily when responding to concerns or issues [99,115]. The availability of information also benefited administrative staff, such as health information managers, who could easily look up mental health information [102]. Improved access to information was also viewed as contributing toward safer and higher quality care [99,102,116]. Boyer et al [103] reported that 74.8% (86/115) of health care professionals interviewed in a psychiatric hospital reported improved access to service user information with an EHR. However, not all information is available on EHRs [112]. Clinicians may have to go through a complex process of identifying what information they need and where they can access that information [104]. Clinicians may also struggle with navigating the EHR because of the amount of information it contains, which is an issue when EHRs do not include search functions [104,106].

Finally, information may be collected for several purposes. The availability of information for one purpose, such as providing care, may not necessarily mean availability for another purpose, such as reporting [100]. Larrison et al [94] found that for community mental health agencies, "capturing data to improve reporting capabilities" was a key motivation for implementing an EHR.

Creates New Structures That Shape Information Collection

This theme reflects the finding that the adoption of the EHRs introduced new structures that shape information collection. These structures standardize and formalize information collection and raise several issues, especially in the mental health context, where unstructured narrative information is used extensively.

Standardized Information

The issue of data standardization arose in several articles, where data fields in the EHR were not suitable for mental health information. Structured fields cannot easily capture the *gray* narrative information common in mental health contexts, and trying to fit data into structured fields can have implications for care [101]. Two common issues were restrictive templates that took away from the narrative format of mental health notes [101,113] and essential templates or data fields missing from the EHR [99,106,112]. Common information collection forms used for mental health care, such as care plans and mental health screening tools, were missing in several EHRs [101,109,113]. When forms were missing, individual clinicians had to decide how to record the information [106]. In some cases, clinicians created standalone tools, such as spreadsheets to collect data. However, this further fractured information in EHRs, unless work was undertaken to integrate the information [112]. EHR formats not being suitable for the mental health context also led to data being captured in other parts of the record, such as free-text boxes or laboratory value areas, which can affect future uses of the data [100,101,109,117]. In addition, the extensive use of free text can make EHRs challenging to navigate [109].

Some of the reasons why standardization did not suit the mental health context included the level of personalization needed in the mental health contexts [111] and that some mental health information is subjective and could be perceived in different ways by different health services [106]. The use of diagnostic codes in an EHR may also create extra work when service users do not clearly fit any one diagnostic code [99]. Specific models of care may also require greater flexibility and personalization of the information collected [111]. An example of this is found in a study on person-centered care planning by Stanhope and Matthews [85,111], who found that standard forms in the EHR were barriers to person-centered care.

Standardization is not necessarily a negative process, and Takian et al [99] found that the standardization of letters sent to people's general practitioners was viewed as beneficial. Clinicians have also recognized the benefits of data management tools to improve the searchability, visibility, and accessibility of information [103,108].

Standardization was also raised as a broader systems issue, where EHRs could not be tailored to specific organizations or settings. This issue was raised in a few studies that adopted commercial EHRs [101,110,111]. In a series of studies from the UK National Health Service (NHS), where uniform EHRs were being adopted, services wanted to tailor the standard solution to their unique needs and the changing priorities of their communities [99,100,107,108].

XSL•FO
RenderX

Informal Versus Formal Documentation

The 8% (2/40) of studies that explored the process that clinicians go through to document information found an element of informality in how mental health information was collected before a specific judgment was made and recorded in the EHR [105,106]. Hardstone et al [105] described how mental health clinicians used informal information practices to develop ideas before they were formalized in the health record. Paper records appeared to enable this informal documentation. In contrast, this provisionality enabled by paper records is limited by EHRs, where the information entered is viewed as a finalized account. Compared with a paper record, recording in an EHR had a greater sense of finality and permanence, which did not align with the informal discussion and sharing of assessments in integrated care settings [105]. Hardstone et al [105] outlined how EHRs may tightly embed rules around who can access records and when, which limits the flexibility to work on notes collaboratively. Zhou et al [106] found that EHRs did not have the functionality to capture provisional information.

Supports Information Sharing and Communication

This theme captures how EHRs supported the components of integrated care, including information sharing and communication among professionals.

Communication Among Service Providers

The specific functions of EHRs may support information sharing and communication among service providers. The functions of EHRs that improve communication include the ability to assign tasks or notes to other clinicians [104], the use of messaging systems [92,101], and shared care plans [112]. These functions that allow clinicians to share information about service users can support the tailoring of care, reduction of unnecessary assessment, and reduction in the number of times service users have to retell the theory story [101]. However, not all EHRs had these functions [112]. There is some evidence that EHRs can improve service users' experience of integrated care. Hu et al [90] found that EHR adoption was significantly associated with improved service user experience for "care transition" and "discharge information" in psychiatric hospitals. Jetelina et al [110] also found a significant improvement in service users' perceptions of integrated care after the implementation of a mental health–specific EHR. However, EHRs that support integrated care may have to be situated in a model of care [85].

Interoperability Between EHRs and Services

Interoperability was raised as an issue across several contexts in the included studies. Several papers acknowledged that integrated EHRs are not always linked with all relevant mental health services [99,100,107,115]. An issue raised in implementing a national EHR in the UK NHS [100,107] was drawing boundaries regarding what services and clinicians can access the EHR. Ser et al [100] outlined how some local community services' information systems were not integrated into the EHR, although these services played a significant role in providing mental health care. Robertson et al [107] also acknowledged that individuals may receive care from many services that are not always contained within a specific geographic region, which an EHR was designed to include. Furthermore, some EHRs lacked interoperability within and among health services [112]. Workarounds for the lack of interoperability identified by Cifuentes et al [112] included printing information from one EHR and scanning it into another EHR or duplicating documentation, which created delays and extra work.

Disrupts Information Management Workflows That Affect the Therapeutic Relationship

This theme explores how EHRs disrupt information practice workflows and raise concerns regarding therapeutic relationships.

Workflow Disruption

Nonalignment of EHRs with workflows was raised in several studies [84,100,103,108,116]. For example, 34.6% (9/26) of psychology training clinics represented in the study by Cellucci et al [116] raised "the difficulty of getting EMR to do what they wanted" as a barrier to implementation. Boyer et al [103] found that 73% (84/115) of interviewed health care professionals in a psychiatric hospital raised the issue of workflow in connection with reduced efficiency, specifically, the challenge of balancing service user care needs and using the EMR. Workflow misalignments led to less time for direct care, which was viewed as affecting the therapeutic relationship [87,100,103].

Sheikth et al [108], Takian et al [99], and Edwards et al [84], in examples of the RiO EHR from the UK NHS, outlined how mental health presentations were complex and varied and required long and detailed assessments. Participants raised that it would not be feasible to try and get people in a crisis setup near a computer so that they could input notes simultaneously [108]. This situation may lead to information having to be inputted later, which could have a broader impact on the operations of the hospital [108]. Participants in the study by Ser et al [100] outlined the challenge of balancing EHR use and supporting people in a crisis, which is common in the mental health context. In clinical therapy, Matthews [101] found that some specific psychological therapies that are more structured may be appropriate for EHR documentation, such as cognitive behavioral therapy.

Matthews [101] and Ser et al [100] found that the EHR interface and design were more medically orientated and designed for contexts in which service users could be treated and discharged and did not need ongoing care. They also found that EHRs missed key mental health functions such as treatment planning and mental health screening. Workarounds were developed to overcome these EHR issues; however, they could be time consuming and require extra work [101]. In comparison, participants in the study by Jung et al [102] who used an EHR specifically designed for mental health contexts commented that they appreciated the EHR being designed for their workflow, including multidisciplinary documentation functions. Administrative staff, including health information managers, valued the ability to make changes to the templates in the EHR where necessary [102]. Similar findings were apparent in the research by Jetelina et al [110], where a mental health–specific add-on to an EPIC brand of the EHR system was evaluated.

XSL•FO
RenderX

The tool improved screening and had good acceptability by clinicians.

The Therapeutic Relationship

The findings regarding the impact of EHRs on therapeutic relationships were mixed. Stewart et al [82] found no significant impact on the therapeutic relationship in a survey of people accessing outpatient psychiatric services where EHRs were used. In interviewing health care professionals at a psychiatric hospital, Boyer et al [103] found that 47% (54/115) were concerned about the triangulation of the therapeutic relationship with the inclusion of an EMR. Interestingly, Matthews [83] found that clinicians rated EHRs as more disruptive to care than service users did. This difference could be explained by the finding that clinicians used a number of strategies to integrate EHR into a session to minimize disruption for service users [83]. Conversely, EHRs have been seen as strengthening the therapeutic relationship by opening the documentation process to service users for discussion and better tailoring care to service users' needs [101].

User Design, Computer Literacy, and the Learning Curve

Several studies have reported that EHRs' complex user interface designs contributed to workflow disruption [99,101,102,111,113,115]. This complexity was related to navigating multiple screens and menus and working with complex templates. Matthews [101] found that clinicians had to navigate various parts of the EHR (screens, menus, and tabs) to record information and that templates did not always follow a structured order that was relevant to the session's progress. Some of these issues may also be specific to the type of clinician. Jung et al [102] found that nurses who had the broadest range of access within the EHR experienced confusion because of the number of modules and the amount of information available to them. Issues with user interface design led to increased time burden for clinicians when documenting information in the EHR [100,101]. Alerts in the EHR were raised as issues in 8% (2/40) of the included studies [102,115]. Some studies reported frequent system crashes or technical glitches such as server issues, which severely affected EHR use and care provision [100,101,111,113]. Participants in the study by Takian et al [99] reported issues logging in and out of systems, especially as legacy systems were running alongside the EHR.

Low computer literacy was raised as a reason why clinicians may find the user interface of the EHR complex [100-102]. Clinicians may also have variable computer skills, specific skills such as typing, and general skills in using technology [88,100-102,107]. For some clinicians, the learning curve can be quite significant [110]. Sheikh et al [108] also found that EHRs may be designed for one type of clinician rather than for many health care professionals and administrative staff using the EHR. Several studies raised the importance of high-quality training to address usability issues [99,102,116].

Challenges Clinician's Management of Sensitive Information

This theme relates to how EHRs raise issues regarding the management of sensitive information and how reducing access to certain parts of the EHR was a common approach to managing issues of sensitive information.

Sensitive Information

Several studies acknowledged that information collected in the mental health context could be particularly sensitive, such as information on traumatic personal events [80,89,93,97,103,113]. EHRs may lead to sensitive information collected by clinicians being more widely available to other clinicians, thus challenging the confidentiality between service users and clinicians [88,100,103]. Several studies explored how specific conditions, including posttraumatic stress disorder [96], substance use [97], mental health diagnoses among people living with HIV [98], and sexual trauma [93], were documented in EHRs. These topics were generally contextualized as sensitive, which affected their documentation. In studies that explored clinicians' documentation practices, an approach clinicians took to sensitive information was generalizing it or *watering it down* [89,100]. Another approach was excluding this information from the EHR or keeping a *shadow record* or paper record for mental health information [89,96]. A finding from the study by Zhou et al [106] points to the subjectivity in clinicians' decisions regarding when to document mental health information.

In some cases, concerns about sensitive information were related to a lack of clarity regarding the legal requirements regarding privacy and confidentiality [113] and the need for further training on these topics [116]. Psychiatric health care professionals in the study by Boyer et al [103] raised the issue of balancing the need to record sensitive information for the provision of care with the risk that it may be used to create a *profile* of service users for other purposes.

Mutual Access to Psychiatric Information

A common indirect way that sensitive information was raised as an issue was by sectioning mental health records in the EHR [80,86,89,97,113]. By sectioning the record, nonpsychiatric clinicians could not access mental health notes or could only access them with a password or if they were willing to *break the glass* and have their access recorded. For example, Bhe et al [81] reported that psychiatrists were given the option of creating two separate notes in the EHR, one accessible by other psychiatric clinicians and one for nonpsychiatric clinicians.

There is evidence that mutual access to psychiatric information supported the provision of mental health care. Bhe et al [81] found that primary care clinicians valued access to psychiatric information as it enabled them to provide care relevant to someone's psychiatric needs, such as by considering the side effects of medication. Mutual access to mental health records may also support care coordination between mental health care and primary health care providers [86,113]. Colaiaco et al [86] found, in practices with a mutual EHR, that 46% (19/41) of reviewed service users' primary care records showed some contact between primary health care and mental health care clinicians compared with only 11% (11/100) in practices with no mutual EHR. Furthermore, 100% (24/24) of the reviewed records in services with a mutual EHR had medication information updated across mental health and primary care

XSL•FO
RenderX

providers' records compared with 57% (31/54) in nonmutual EHR services.

This study does not seek to consider the clinical implications of EHRs. However, we would be remiss not to mention a finding from the study by Kozubal et al [80] that there was a significant relationship between increased accessibility (nonpsychiatric clinicians' ability to access psychiatric records) and reduced readmission rates.

Raises Legal Concerns for Clinicians Regarding Their Information Responsibilities

The final theme regarding legal issues, particularly those related to privacy and mental health laws, appeared in far fewer studies than we had anticipated. There was little congruence among the references to legal concerns, with a variety of different concerns raised across the included studies. Reitz et al [113] found that the use of EHRs raised concerns about compliance with relevant information privacy laws. Ser et al [100] found that clinicians expressed concerns about whether EHRs aligned with their requirements under relevant mental health legislations. In the study by Jung et al [102], administrative staff, such as health information managers, outlined how the EHR supported compliance with relevant regulations by reducing the reporting burden. Clinical staff also reported wanting alerts relevant to their legal requirements when people were being treated under the relevant mental health laws [102]. Participants in the survey by Cellucci et al [116], representing psychology training clinics, identified the need for training on ethical issues, confidentiality, and security standards. Participants in the study by Matthews [101] outlined how state regulations and standards shaped the design and use of EHRs.

Discussion

Principal Findings

This scoping review aimed to explore how EHRs in the mental health context affected the information practices of health care professionals and how these changes affected other aspects of care. Issues relevant to the mental health context, such as the management of sensitive information, data standardization, and therapeutic relationships, were also explored. We found that EHRs can improve some information practices but need to be designed appropriately for specific workflows and information types in the mental health context. Beyond the design of EHRs, the redesign of health service workflows and clinician training may be needed to ensure that EHRs can be used effectively in the mental health context. Information collected in the mental health context is considered more sensitive than other types of clinical information. Greater guidance may be needed regarding how sensitive information is managed in EHRs to ensure that it is documented and used appropriately. In the following sections, we consider how the findings of this review link back to the broader literature on EHRs.

The documentation of clinical information is a critical information practice that informs current and future care [119-121]. The findings of this review point to improvements in the relative quantity of the information documented when using an EHR compared with paper records. However, information was still missing from EHRs, which may affect future care. Furthermore, a common issue for clinicians was the inflexibility of the fields in EHRs and the time required to input data. This issue may be partly because of the greater coupling of policy and process with tools for documentation, such as templates. Mamykina et al [121] has raised this focus on templates and structures in EHRs as an outcome of viewing clinical documentation as a *composition* activity. However, through a time-and-motion study, they found that clinical documentation was a synthesis activity involving clinicians accessing several informal and formal information sources that they synthesized into clinical documents. This reflects the finding from this review that informal documentation is a necessary precursor to formal documentation and contributes to the synthesis of the final documentation. Mamykina et al [121] argued that tools for composition, such as templates, differ from tools for synthesis, which should promote access to various information sources, such as informal notes that previously could be written and edited within the paper document. This finding may explain why certain information is missed in the structured documentation in EHRs, although it was available in other free-text sections.

The focus on the standardization and the formalization of documentation exposed a critical tension between current approaches to health informatics and contemporary mental health care. An objective of EHRs is the standardization of health information to allow for health information exchange and data analytics [122,123]. In comparison, mental health care involves the documentation of a large amount of narrative information, much of which resists standardization [16,51]. An increasing focus on recovery models of mental health care that prioritize service user–defined measures and outcomes may create further tensions with standardized data collection [124]. Concerns have also been raised about EHRs impeding clinicians' ability to understand a service user's entire story [125]. These issues were discussed in 1998 regarding the need for an informatics framework specific to mental health [126]. Future research and EHR design need to establish which standardized information is relevant for the mental health context and how best to present narrative information to capture service users' stories.

The issue of standardization found in this review is not unique to mental health in that paper records, in general, provide more opportunities than EHRs for recording narrative information [127]. The many benefits promised by EHRs in terms of decision support, streamlined reporting, and supporting research are premised on the need for structured data entry [24]. However, narrative information is highly valued by clinicians. This may reflect why clinicians used narrative information, even when structured fields were available. Our findings and research in other contexts indicate that clinicians prefer documentation methods that align with their workflows and allow them to record more details about clinical encounters [24,128-130]. Narrative information provides opportunities for clinicians to convey information such as uncertainty, unique aspects of cases, and nuances in the service user's appearance, which is not supported by structured documentation [24]. A potential solution to the tension between unstructured and structured

XSL•FO
RenderX

documentation is the application of software to unstructured clinical notes that can extract relevant data into structured fields. For example, natural language processing could be applied to free-text narratives to fill structured EHR fields [24,127].

The management of sensitive information was raised as a key concern in the adoption of EHRs. The definitions of and overlap between sensitive information and mental health information are unclear. The National Committee on Vital and Health Statistics [131] outlined the complexity of defining mental health information in that it can be collected in a variety of clinical settings and may be scattered throughout a person's health record. Data about physical health, collected in mental health settings, may also be considered mental health information. However, there appears to be a subsection of mental health information classified as *sensitive* for several reasons, such as the stigma related to certain diagnoses. There is also a relationship between standardization and sensitive information, with some studies in this review finding that sensitive information may be captured in free-text notes but not in standardized fields. Perhaps free text provides more nuances in documenting this type of information. For example, Cairns et al [132], in a study of social workers using a shared record, found that they had concerns about reporting subjective information that other people could wrongly interpret. One of the potential issues with incorrect or vague documentation in the mental health context is that it could feed into incorrect risk assessments [133]. Risk assessments in mental health can have significant implications for people's health outcomes and their human rights if a risk assessment leads them to be involuntarily treated [133].

Shared decision-making has become a key approach for promoting autonomy in health care decision-making, especially in the mental health context [134]. This can be seen in the practices of clinicians inviting service users to be involved in deciding what information to document in their health record, which is known as collaborative documentation [91]. Inviting service users to participate in decisions about what information goes into their EHR and how to document sensitive information could help address concerns that clinicians might have about privacy or stigma. Pisciotta et al [135] found that clinicians and service users in mental health settings avoided discussing notes because they worried about each other's responses. Pisciotta et al [135] also found that service users want clinicians to be open to discussing what is written about them and have opportunities to collaborate in documenting information. Collaborative documentation may also address concerns about the therapeutic alliance if workflows are redesigned to accommodate EHRs and service users. Maniss and Pruit [136] outlined how collaborative documentation involves clinicians documenting service user information alongside service users and creating opportunities for their input and approval. However, as was found in the included studies, the current EHR design is not aligned with the complexity of some mental health contexts where service users may arrive in a crisis. Thus, the adoption of collaborative documentation may need to happen alongside other service changes to ensure that EHRs can be more easily integrated into service users' care.

The findings related to the relationship between information practices and therapeutic relationships require more research, especially from the service user perspective. It has been suggested that most information practices are invisible to service users [137] unless there are active efforts to make them visible. However, these practices and how they are shaped using EHRs will affect service users' experience of care through impacts on the therapeutic relationship or the time available for direct care. Much research has focused on service users' perspectives concerning the privacy and confidentiality of EHRs [12]; however, the actual impact on the experience of care has received limited attention. There is a growing body of evidence exploring the role of computers in clinical encounters, which may capture some of these experiences [138,139]. The impact of computers on the interaction between clinicians and service users can be shaped by factors such as the clinician's skill in using the computer and the way clinicians embed computers in their practice [140]. Findings from the study by Pearce et al [141] showed that computers had become part of a *triadic relationship* with clinicians and service users, which is not necessarily a negative outcome. Future work should explore how EHRs as sociotechnical systems affect the care provided and service users' experience of these impacts.

Comparison With Prior Work

Overview

There are several reviews related to different elements of EHRs, which generally support the findings of this review. In a systematic review of the impact of EHRs on documentation time, Baumann et al [23] found that EHRs were associated with increased documentation time for hospital staff. The interaction between service users and clinicians was also raised as potentially threatened by the use of EHRs [142]. Workflow issues were also identified by Gephart et al [143] in a systematic review of nurses' experiences of EHR. They found that EHRs created unexpected changes in the accepted workflows. Strudwick and Eyasu [144], in a review of the literature on EHRs used by nurses in mental health settings, also identified the unique nature of the mental health context. They found that nurses were conscious of the privacy and confidentiality risks posed by the ease of access enabled by EHRs. A recent scoping review on EMR implementations in mental health settings by Zurynski et al [30], which also included studies in children and adolescents and several review studies not specific to the mental health context, also found issues with documentation, workflows, and usability.

The issue of usability that was raised in this study has been confirmed by previous reviews exploring navigation in EHRs [143,145,146]. Roman et al [145] found that navigation between EHR screens was a regularly identified usability issue that could be addressed through shortcuts, dashboards, and integration of information into single screens. Training has also been found to enable the acceptance and use of EHRs [146,147]. McGinn et al [142], in a systematic review of barriers to and facilitators of EHR implementation, also found that usability could be both a barrier to and a facilitator of EHR use.

An increasing number of studies have identified new secondary uses for the data collected in EHRs [148]. These secondary uses

XSL•FO
RenderX

include applications in psychiatric phenotyping [149] and methods for predicting suicidal behavior [150]. The potential impact of this secondary data use makes it increasingly urgent to address the issues raised in this study. Secondary data use in the mental health context requires further ethical consideration, especially as new data sources are being introduced into the health care system, such as data from wearables [151,152].

Relevance of Findings for EHR Designers

One of the key issues identified in this review was that EHRs were not appropriately designed for the mental health context. Thus, we will target our recommendations for those who design and develop EHRs.

Designers must ensure that they understand clinicians' information practices in the mental health context. There are examples of EHR usability frameworks such as the TURF (task, user, representation, and function) framework [153], which can guide EHR design. A key point raised by the TURF framework is the need to understand the complexity of a task independent of how it is implemented in a specific setting. Our review found that many EHRs were not designed to address the complexity of the mental health context. This issue could be because of designers and developers not understanding the essential elements of certain tasks and how these should be represented in the design of the EHR. Our review also found that many EHRs are missing data fields relevant to mental health and provide limited ways of managing narrative data. Thus, improving the customizability of EHR workflows may be useful. Alternatively, several preset workflows could be provided for different types of service users or clinical contexts. The study by Jetelina et al [110] provides an example in which an add-on for an EPIC EHR was developed containing specific features for the mental health context. Designers should also consider the computer literacy of their end users and what relevant training and support may be needed.

The time burden experienced by clinicians when documenting information in EHRs raises questions about the systems' user experience design. This time burden is not just an issue in the mental health context, with O'Brien et al [154] describing the broader issue as *death by data entry.* This is a critical issue for clinicians and is associated with increased odds of burnout [155,156]. This could be addressed in several ways, including through research, policy initiatives, and design methods [157]. Our findings suggest that further research is needed for workflows in the mental health context and how EHR functions can support rather than disrupt these workflows. Addressing this issue should also lead to greater end user involvement in designing, developing, and implementing EHRs in the mental health context [158]. Improving clinician training may also support the use of EHRs [99,100].

Relevance of Findings for Clinicians

A key issue found in this review was the poor documentation of mental health information in EHRs. Missing information is detrimental to both the care of service users and clinicians' work. It appears that there is a perception among clinicians that mental health information, being particularly sensitive, should be documented differently from other information. We would advise clinicians to consider approaches such as collaborative documentation in which service users are involved in discussions about what to document. If there is doubt about how to word certain sensitive information, clinicians should ask the service user and consider the implications for future clinical encounters and the service user's experience if certain information is missing or misinterpreted.

Relevance of Findings for Health Service Managers and Health Policy Makers

From this initial evidence on EHRs in the mental health context, it would be advisable for health service managers to scope their options when adopting an EHR. Services should start by identifying their information and workflow needs before choosing an EHR. Some EHRs designed specifically for the mental health context are more appropriate than generic EHRs. Otherwise, specific add-ons that meet the workflow and information needs of the mental health context may be considered. Furthermore, well-executed training is necessary to ensure that clinicians have appropriate computer skills to manage the complex user interfaces that some EHRs present.

We would advise policy makers to support the adoption of EHRs only when they are designed for local contexts. In Australia, the Victorian Royal Commission into Mental Health Services has recently recommended that information systems should be used to improve care in the mental health context [159]. We would advise that further research is needed to identify the mechanisms by which EHRs will lead to the assumed outcomes and any barriers or structural issues to achieving these assumed outcomes.

Relevance of Findings for Service Users

It was concerning that there was minimal involvement of service users in the included studies. The issues identified in this review will have implications for service users. These impacts may be related to disrupted workflows or sensitive information being recorded incorrectly. Many service user groups are taking great interest in the digitization of the health system, and we would encourage them to continue this involvement, especially with a focus on EHR development.

Relevance of Findings for Future Research

Future researchers should report on the types and functions of the EHRs they are studying. This would enable greater comparison between contexts. Adopting a standard approach to describing EHRs such as the Health Care Information and Management Systems Society's [160] EMR Adoption Model may support comparison across studies. Furthermore, more details about the setting of the research should be provided. Health information technology is a global business, and companies providing EHRs to the United States also provide them to other countries. Providing more details about the setting of implementation and the type of EHR would support evidence synthesis that other jurisdictions can also rely on.

Future research should also include service user perspectives on EHRs and information practices. Researchers should consider adopting co-design or participatory methods to involve service users in research about EHRs. It would also be advisable to

XSL•FO
RenderX

involve peer workers within health services in the design of EHRs as they may see how these EHRs have been used in practice. Researchers should also involve more health information managers and other administrative staff. These stakeholders play a critical role in supporting the correct management of information in health care settings.

The field of research on EHRs in the mental health context is still at a low stage of maturity, and this, in part, reflects the maturity of EHR use in the mental health context. Future research should include high-quality evaluations of EHRs in the mental health context for both implementation and sustained use. This research will pave the way for systematic reviews that can provide insights into how EHRs affect processes and clinical outcomes in the mental health context. We would also recommend further studies on the usability of EHRs or that usability analysis be included in other study designs.

Notably, we could not conduct a temporal analysis of the included papers. Recent decades have seen considerable advancements in the fields of health informatics and digital health [161]. It would make sense that these advances should be reflected in the included papers. We might expect to see improvements in interoperability because of the increasing adoption of solutions such as the Fast Health care Interoperability Standard [162]. We may also expect to see improvements in the documentation of standard information using clinical terminologies such as the Systematized Nomenclature of Medicine [163]. However, these advances were not discussed in the included papers. We can speculate why this was the case. It might be that these innovations have not penetrated the mental health context or affected the issues we have identified. However, what is needed in future research is a greater focus on the technical aspects of digital health research. Future studies should aim to report the technical aspects of EHRs in practice to enable greater visibility of how EHR innovations penetrate real-world applications.

Finally, a further piece of research that should be considered is how digital health or information system theory can accommodate the findings of this review and others related to the use of EHRs. A few of the included studies drew upon the theory in their work; however, more work could be conducted to extend this work. We have reflected that many of our findings could be considered using an *Activity Theory* lens, and we would welcome discussions and collaborations to further this thinking.

Limitations

This scoping review is limited, in part, because of the nature of the field. The combination of no standard EHR definition and poor reporting of the systems used in the included studies has made it difficult to assess how specific themes related to specific types of EHRs. This review examines information practices, which is one of the many potential research topics that could be addressed in this space. Other studies should examine clinical outcomes. We expected to find more studies reporting on legal and ethical concerns, and in hindsight, a more tailored search may have been needed. There was limited information on the technical aspects (such as interoperability standards) of the EHRs used in the included studies. This limited information affected our ability to comment on whether the technical elements of the EHR contributed to our findings. The United States' focus of the included studies also limits the applicability of the findings to other jurisdictions, especially those related to health system structure and culture.

Conclusions

EHRs in the mental health contexts have been slow to materialize. This review found that EHRs in the mental health context affect clinicians' information practices, which have implications for how care is provided. The core of mental health services is the therapeutic relationship, which requires a unique workflow that is currently not supported by many EHRs. In addition, because of the narrative nature of mental health care, the standardized data underpinning many EHRs may not align with the information needs and practices of the mental health context. Finally, although health information is recognized as personal information, mental health information is seen as especially sensitive for several reasons. This understanding of mental health information may lead to underreporting, generalizing, or watering down certain details when documenting in EHRs. EHRs need the capacity to support information sharing in a nuanced way to manage sensitivity and stigma in the mental health context. Future research should involve service users to explore how the impact of EHRs on information practices also affects their experience of care.

Acknowledgments

TCK was supported by the Australian Government Research Training Program Stipend and Research Training Program Fee-Offset Scholarship through The University of Melbourne.

The authors wish to thank the 3 anonymous reviewers who provided extensive and thought-provoking feedback on our original submission.

Authors' Contributions

TCK performed the conceptualization, writing of the original draft, and investigation of this paper. SG, MP, and KG each reviewed, edited, and provided supervision.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Included studies.
[DOCX File , 28 KB - jmir_v24i5e30405_app1.docx]

Multimedia Appendix 2
Quotes to support themes.
[DOCX File , 23 KB - jmir_v24i5e30405_app2.docx]

References

1. Jha AK. The promise of electronic records: around the corner or down the road? JAMA 2011;306(8):880-881. [doi: 10.1001/jama.2011.1219] [Medline: 21862751]
2. Brown SA, Jouni H, Marroush TS, Kullo IJ. Disclosing genetic risk for coronary heart disease: attitudes toward personal information in health records. Am J Prev Med 2017;52(4):499-506 [FREE Full text] [doi: 10.1016/j.amepre.2016.11.005] [Medline: 28062272]
3. Scruth EA. Quality nursing documentation in the medical record. Clin Nurse Spec 2014;28(6):312-314. [doi: 10.1097/NUR.0000000000000085] [Medline: 25295557]
4. Richter P, Gand K, Esswein W. Lifetime electronic health records as an enabler for integrated care. In: 49th Hawaii International Conference on System Sciences. 2016 Presented at: HICSS '16; January 5-8, 2016; Koloa, HI p. 3199-3207. [doi: 10.1109/hicss.2016.401]
5. Integrated care models: an overview. World Health Organization. 2016. URL: http://www.euro.who.int/__data/assets/pdf_file/0005/322475/Integrated-care-models-overview.pdf [accessed 2021-01-10]
6. Nicholson C, Jackson CL, Marley JE. Best-practice integrated health care governance - applying evidence to Australia's health reform agenda. Med J Aust 2014;201(3 Suppl):S64-S66. [doi: 10.5694/mja14.00310] [Medline: 25047885]
7. Kariotis T, Prictor M, Gray K, Chang S. Mind the gap: information sharing between health, mental health and social care services. Stud Health Technol Inform 2019;266:101-107. [doi: 10.3233/SHTI190780] [Medline: 31397309]
8. Feeney L, Moran P. Information sharing and risk: a survey of psychiatrists. Ir J Psychol Med 2007;24(3):94-98. [doi: 10.1017/S0790966700010387] [Medline: 30290488]
9. Kariotis TC, Harris KM. Clinician perceptions of My Health Record in mental health care: medication management and sharing mental health information. Aust J Prim Health 2019;25(1):66-71. [doi: 10.1071/PY17181] [Medline: 30636668]
10. Jones IR, Ahmed N, Catty J, McLaren S, Rose D, Wykes T, Echo group. Illness careers and continuity of care in mental health services: a qualitative study of service users and carers. Soc Sci Med 2009;69(4):632-639. [doi: 10.1016/j.socscimed.2009.06.015] [Medline: 19577834]
11. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. Healthc Manage Forum 2017;30(2):111-116 [FREE Full text] [doi: 10.1177/0840470416679413] [Medline: 28929889]
12. Shen N, Sequeira L, Silver MP, Carter-Langford A, Strauss J, Wiljer D. Patient privacy perspectives on health information exchange in a mental health context: qualitative study. JMIR Ment Health 2019;6(11):e13306 [FREE Full text] [doi: 10.2196/13306] [Medline: 31719029]
13. Myllärniemi J, Laihonen H, Karppinen H, Seppänen K. Knowledge management practices in healthcare services. Meas Bus Excell 2012;16(4):54-65. [doi: 10.1108/13683041211276447]
14. Fourie I. Learning from research on the information behaviour of healthcare professionals: a review of the literature 2004-2008 with a focus on emotion. Health Info Libr J 2009;26(3):171-186 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00860.x] [Medline: 19712210]
15. Coiera E. Guide to health informatics. 3rd edition. Boca Raton, FL: CRC Press; 2015.
16. Lawn S, Zabeen S, Rowlands N, Picot S. Hidden care: revelations of a case-note audit of physical health care in a community mental health service. Int J Ment Health Nurs 2018;27(6):1742-1755. [doi: 10.1111/inm.12479] [Medline: 29797643]
17. Alexander MJ, Siegel C, Murtaugh C. Automating the psychiatric record for care review purposes: a feasibility analysis. Comput Hum Services 1985;1(4):1-16. [doi: 10.1300/j407v01n04_01]
18. Ventres W, Kooienga S, Vuckovic N, Marlin R, Nygren P, Stewart V. Physicians, patients, and the electronic health record: an ethnographic analysis. Ann Fam Med 2006;4(2):124-131 [FREE Full text] [doi: 10.1370/afm.425] [Medline: 16569715]
19. Makoul G, Curry RH, Tang PC. The use of electronic medical records: communication patterns in outpatient encounters. J Am Med Inform Assoc 2001;8(6):610-615 [FREE Full text] [doi: 10.1136/jamia.2001.0080610] [Medline: 11687567]
20. Patel VL, Kushniruk AW, Yang S, Yale JF. Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. J Am Med Inform Assoc 2000;7(6):569-585 [FREE Full text] [doi: 10.1136/jamia.2000.0070569] [Medline: 11062231]
21. Wrenn JO, Stein DM, Bakken S, Stetson PD. Quantifying clinical narrative redundancy in an electronic health record. J Am Med Inform Assoc 2010;17(1):49-53 [FREE Full text] [doi: 10.1197/jamia.M3390] [Medline: 20064801]

XSL•FO
RenderX

22. Campanella P, Lovato E, Marone C, Fallacara L, Mancuso A, Ricciardi W, et al. The impact of electronic health records on healthcare quality: a systematic review and meta-analysis. Eur J Public Health 2016;26(1):60-64. [doi: 10.1093/eurpub/ckv122] [Medline: 26136462]
23. Baumann LA, Baker J, Elshaug AG. The impact of electronic health record systems on clinical documentation times: a systematic review. Health Policy 2018;122(8):827-836. [doi: 10.1016/j.healthpol.2018.05.014] [Medline: 29895467]
24. Rosenbloom ST, Denny JC, Xu H, Lorenzi N, Stead WW, Johnson KB. Data from clinical notes: a perspective on the tension between structure and flexible documentation. J Am Med Inform Assoc 2011;18(2):181-186 [FREE Full text] [doi: 10.1136/jamia.2010.007237] [Medline: 21233086]
25. Walsh SH. The clinician's perspective on electronic health records and how they can affect patient care. BMJ 2004;328(7449):1184-1187 [FREE Full text] [doi: 10.1136/bmj.328.7449.1184] [Medline: 15142929]
26. Busch AB, Bates DW, Rauch SL. Improving adoption of EHRs in psychiatric care. N Engl J Med 2018;378(18):1665-1667 [FREE Full text] [doi: 10.1056/NEJMp1800465] [Medline: 29719176]
27. The eHealth readiness of Australia's allied health sector - final report. Department of Health. 2011. URL: http://www.health.gov.au/internet/publications/publishing.nsf/Content/ehealth-readiness-allied-toc [accessed 2019-03-01]
28. Krist AH, Beasley JW, Crosson JC, Kibbe DC, Klinkman MS, Lehmann CU, et al. Electronic health record functionality needed to better support primary care. J Am Med Inform Assoc 2014;21(5):764-771 [FREE Full text] [doi: 10.1136/amiajnl-2013-002229] [Medline: 24431335]
29. Wykes T, Brown M. Over promised, over-sold and underperforming? - e-health in mental health. J Ment Health 2016;25(1):1-4 [FREE Full text] [doi: 10.3109/09638237.2015.1124406] [Medline: 26732733]
30. Zurynski Y, Ellis LA, Tong HL, Laranjo L, Clay-Williams R, Testa L, et al. Implementation of electronic medical records in mental health settings: scoping review. JMIR Ment Health 2021;8(9):e30564 [FREE Full text] [doi: 10.2196/30564] [Medline: 34491208]
31. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. J Med Syst 2016;40(12):252 [FREE Full text] [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]
32. Wibe T, Ekstedt M, Hellesø R. Information practices of health care professionals related to patient discharge from hospital. Inform Health Soc Care 2015;40(3):198-209. [doi: 10.3109/17538157.2013.879150] [Medline: 24475936]
33. Reddy MC, Jansen BJ. A model for understanding collaborative information behavior in context: a study of two healthcare teams. Inf Process Manag 2008;44(1):256-273. [doi: 10.1016/j.ipm.2006.12.010]
34. French RL, Williamson K. The information practices of welfare workers: conceptualising and modelling information bricolage. J Doc 2016;72(4):737-754. [doi: 10.1108/jdoc-08-2015-0100]
35. Østensen E, Bragstad LK, Hardiker NR, Hellesø R. Nurses' information practice in municipal health care-a web-like landscape. J Clin Nurs 2019;28(13-14):2706-2716. [doi: 10.1111/jocn.14873] [Medline: 30938870]
36. Lea NC, Nicholls J. Are patient relationships the driver for information governance? Br J Gen Pract 2016;66(648):342-343 [FREE Full text] [doi: 10.3399/bjgp16X685753] [Medline: 27364653]
37. Fitch SK. Trust, information-sharing and the doctor-patient relationship: a multi-method, empirical-ethics study of New Zealand general practice. The University of Auckland. 2017. URL: https://researchspace.auckland.ac.nz/bitstream/handle/2292/33558/whole.pdf?sequence=2 [accessed 2019-03-20]
38. Firn J, Preston N, Walshe C. Ward social workers' views of what facilitates or hinders collaboration with specialist palliative care team social workers: a grounded theory. BMC Palliat Care 2017;17(1):7 [FREE Full text] [doi: 10.1186/s12904-017-0214-z] [Medline: 28705157]
39. Rothstein MA. The Hippocratic bargain and health information technology. J Law Med Ethics 2010;38(1):7-13 [FREE Full text] [doi: 10.1111/j.1748-720X.2010.00460.x] [Medline: 20446978]
40. Kariotis T, Prictor M, Chang S, Gray K. Evaluating the contextual integrity of Australia's my health record. Stud Health Technol Inform 2019;265:213-218. [doi: 10.3233/SHTI190166] [Medline: 31431601]
41. Kämpf A, McSherry BM, Ogloff JR, Rothschild AD. Confidentiality for mental health professionals: a guide to ethical and legal principles. Bowen Hills, Australia: Australian Academic Press; 2009.
42. Ho HD, Satur J, Meldrum R. Perceptions of oral health by those living with mental illnesses in the Victorian Community - the consumer's perspective. Int J Dent Hyg 2018;16(2):e10-e16. [doi: 10.1111/idh.12278] [Medline: 28370974]
43. Krotofil J, McPherson P, Killaspy H. Service user experiences of specialist mental health supported accommodation: a systematic review of qualitative studies and narrative synthesis. Health Soc Care Community 2018;26(6):787-800. [doi: 10.1111/hsc.12570] [Medline: 29609195]
44. Nash M. Mental health service users' experiences of diabetes care by mental health nurses: an exploratory study. J Psychiatr Ment Health Nurs 2014;21(8):715-723. [doi: 10.1111/jpm.12140] [Medline: 24548452]
45. Aviram RB, Brodsky BS, Stanley B. Borderline personality disorder, stigma, and treatment implications. Harv Rev Psychiatry 2006;14(5):249-256. [doi: 10.1080/10673220600975121] [Medline: 16990170]
46. Grando A, Ivanova J, Hiestand M, Soni H, Murcko A, Saks M, et al. Mental health professional perspectives on health data sharing: mixed methods study. Health Informatics J 2020;26(3):2067-2082 [FREE Full text] [doi: 10.1177/1460458219893848] [Medline: 31928295]

XSL•FO
RenderX

47. Shank N. Behavioral health providers' beliefs about health information exchange: a statewide survey. J Am Med Inform Assoc 2012;19(4):562-569 [FREE Full text] [doi: 10.1136/amiajnl-2011-000374] [Medline: 22184253]
48. McGregor B, Mack D, Wrenn G, Shim RS, Holden K, Satcher D. Improving service coordination and reducing mental health disparities through adoption of electronic health records. Psychiatr Serv 2015;66(9):985-987 [FREE Full text] [doi: 10.1176/appi.ps.201400095] [Medline: 25975885]
49. Triplett P. Psychiatry and the meaningful use of electronic health records. Perspect Biol Med 2013;56(3):407-421. [doi: 10.1353/pbm.2013.0028] [Medline: 24375121]
50. Greene J. Behavioral health data in the electronic health record: privacy concerns slow sharing. Ann Emerg Med 2013;62(4):19A-21A. [doi: 10.1016/j.annemergmed.2013.08.003] [Medline: 24180009]
51. Kobus AM, Harman JS, Do HD, Garvin RD. Challenges to depression care documentation in an EHR. Fam Med 2013;45(4):268-271. [Medline: 23553091]
52. Ranallo PA, Kilbourne AM, Whatley AS, Pincus HA. Behavioral health information technology: from chaos to clarity. Health Aff (Millwood) 2016;35(6):1106-1113. [doi: 10.1377/hlthaff.2016.0013] [Medline: 27269029]
53. Tai B, Boyle M, Ghitza U, Kaplan RM, Clark HW, Gersing K. Meaningful use of electronic behavioral health data in primary health care. Sci Transl Med 2012;4(119):119mr3. [doi: 10.1126/scitranslmed.3003324] [Medline: 22301552]
54. Knowles P. Collaborative communication between psychologists and primary care providers. J Clin Psychol Med Settings 2009;16(1):72-76. [doi: 10.1007/s10880-009-9151-1] [Medline: 19234866]
55. Archer C, Kessler D, Wiles N, Turner K. GPs' and patients' views on the value of diagnosing anxiety disorders in primary care: a qualitative interview study. Br J Gen Pract 2021;71(707):e450-e457 [FREE Full text] [doi: 10.3399/BJGP.2020.0959] [Medline: 33824158]
56. Cresswell K, Morrison Z, Kalra D, Sheikh A. "There are too many, but never enough": qualitative case study investigating routine coding of clinical information in depression. PLoS One 2012;7(8):e43831 [FREE Full text] [doi: 10.1371/journal.pone.0043831] [Medline: 22937106]
57. Ford E, Campion A, Chamles DA, Habash-Bailey H, Cooper M. "You don't immediately stick a label on them": a qualitative study of influences on general practitioners' recording of anxiety disorders. BMJ Open 2016;6(6):e010746 [FREE Full text] [doi: 10.1136/bmjopen-2015-010746] [Medline: 27338879]
58. Rosen DC, Nakash O, Alegría M. The impact of computer use on therapeutic alliance and continuance in care during the mental health intake. Psychotherapy (Chic) 2016;53(1):117-123. [doi: 10.1037/pst0000022] [Medline: 26214322]
59. Plovnick RM. The progression of electronic health records and implications for psychiatry. Am J Psychiatry 2010;167(5):498-500. [doi: 10.1176/appi.ajp.2009.09101473] [Medline: 20439396]
60. Gilburt H, Rose D, Slade M. The importance of relationships in mental health care: a qualitative study of service users' experiences of psychiatric hospital admission in the UK. BMC Health Serv Res 2008;8:92 [FREE Full text] [doi: 10.1186/1472-6963-8-92] [Medline: 18439254]
61. Priebe S, Mccabe R. Therapeutic relationships in psychiatry: the basis of therapy or therapy in itself? Int Rev Psychiatry 2008;20(6):521-526. [doi: 10.1080/09540260802565257] [Medline: 19085408]
62. Duffy WJ, Kharasch MS, Du H. Point of care documentation impact on the nurse-patient interaction. Nurs Adm Q 2010;34(1):E1-10. [doi: 10.1097/NAQ.0b013e3181c95ec4] [Medline: 20023554]
63. Shank N, Willborn E, Pytlikzillig L, Noel H. Electronic health records: eliciting behavioral health providers' beliefs. Community Ment Health J 2012;48(2):249-254. [doi: 10.1007/s10597-011-9409-6] [Medline: 21573771]
64. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
65. Lee SJ, Crowther E, Keating C, Kulkarni J. What is needed to deliver collaborative care to address comorbidity more effectively for adults with a severe mental illness? Aust N Z J Psychiatry 2013;47(4):333-346. [doi: 10.1177/0004867412463975] [Medline: 23076548]
66. 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] [Medline: 30453902]
67. 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]
68. 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] [Medline: 30178033]
69. Constitution of the World Health Organization. World Health Organization. 2020. URL: https://apps.who.int/gb/bd/pdf_files/BD_49th-en.pdf#page=6 [accessed 2019-02-10]
70. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008;77(5):291-304. [doi: 10.1016/j.ijmedinf.2007.09.001] [Medline: 17951106]

XSL•FO
RenderX

71. Garets D, David M. Electronic medical records vs. electronic health records: yes, there is a difference. HIMSS Analytics. 2006. URL: https://www.aao.org/asset.axd?id=8e9b1f20-0ed6-4d2b-92f8-e28fbaf378ec&t=634962799822530000 [accessed 2019-02-20]
72. Glossary of acronyms and terms commonly used in informatics. American Medical Informatics Association. 2013. URL: https://www.amia.org/glossary [accessed 2021-02-01]
73. Zhang T, Shen N, Booth R, LaChance J, Jackson B, Strudwick G. Supporting the use of patient portals in mental health settings: a scoping review. Inform Health Soc Care 2022;47(1):62-79. [doi: 10.1080/17538157.2021.1929998] [Medline: 34032528]
74. Moorman PW, van der Lei J. An inventory of publications on electronic patient records. Methods Inf Med 1999;38(4-5):294-297. [Medline: 10805016]
75. Lakshman R, Ogilvie D, Ong KK. Mothers' experiences of bottle-feeding: a systematic review of qualitative and quantitative studies. Arch Dis Child 2009;94(8):596-601 [FREE Full text] [doi: 10.1136/adc.2008.151910] [Medline: 19602520]
76. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
77. Barnett-Page E, Thomas J. Methods for the synthesis of qualitative research: a critical review. BMC Med Res Methodol 2009;9:59 [FREE Full text] [doi: 10.1186/1471-2288-9-59] [Medline: 19671152]
78. Meredith J. Electronic patient record evaluation in community mental health. Inform Prim Care 2009;17(4):209-213 [FREE Full text] [doi: 10.14236/jhi.v17i4.739] [Medline: 20359398]
79. Riahi S, Fischler I, Stuckey MI, Klassen PE, Chen J. The value of electronic medical record implementation in mental health care: a case study. JMIR Med Inform 2017;5(1):e1 [FREE Full text] [doi: 10.2196/medinform.6512] [Medline: 28057607]
80. Kozubal DE, Samus QM, Bakare AA, Trecker CC, Wong HW, Guo H, et al. Separate may not be equal: a preliminary investigation of clinical correlates of electronic psychiatric record accessibility in academic medical centers. Int J Med Inform 2013;82(4):260-267 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.11.007] [Medline: 23266060]
81. Bhe E, Summers S, Pakyurek M, Soulier M, Ferranti J. Outpatient psychiatric documentation use by primary care physicians following de-sensitization in the electronic medical record. Acad Psychiatry 2015;39(6):639-644. [doi: 10.1007/s40596-014-0201-0] [Medline: 25069956]
82. Stewart RF, Kroth PJ, Schuyler M, Bailey R. Do electronic health records affect the patient-psychiatrist relationship? A before & after study of psychiatric outpatients. BMC Psychiatry 2010;10:3 [FREE Full text] [doi: 10.1186/1471-244X-10-3] [Medline: 20064210]
83. Matthews EB. Whose system is it anyway? How clients and providers evaluate the impact of session EHR use on communication in mental health treatment. Community Ment Health J 2020;56(3):540-548. [doi: 10.1007/s10597-019-00511-8] [Medline: 31802323]
84. Edwards K, Chiweda D, Oyinka A, McKay C, Wiles D. Assessing the value of electronic records. Nurs Times 2011;107(40):12-14. [Medline: 22132486]
85. Matthews EB, Stanhope V. A help or a hindrance: the role of electronic health records in implementing new practices in community mental health clinics. Psychiatr Serv 2020;71(10):1065-1068 [FREE Full text] [doi: 10.1176/appi.ps.201900470] [Medline: 32838681]
86. Colaiaco B, Roth CP, Ganz DA, Hanson M, Smith P, Wenger NS. Continuity of information between mental health and primary care providers after a mental health consultation. Psychiatr Serv 2018;69(10):1081-1086. [doi: 10.1176/appi.ps.201800025] [Medline: 30041587]
87. Tsai J, Bond G. A comparison of electronic records to paper records in mental health centers. Int J Qual Health Care 2008;20(2):136-143. [doi: 10.1093/intqhc/mzm064] [Medline: 18079127]
88. Xiao AQ, Acosta FX. Implementation and impact of psychiatric electronic medical records in a public medical center. Perspect Health Inf Manag 2016;13(Fall):1e [FREE Full text] [Medline: 27843422]
89. Salomon RM, Blackford JU, Rosenbloom ST, Seidel S, Clayton EW, Dilts DM, et al. Openness of patients' reporting with use of electronic records: psychiatric clinicians' views. J Am Med Inform Assoc 2010;17(1):54-60 [FREE Full text] [doi: 10.1197/jamia.M3341] [Medline: 20064802]
90. Hu X, Qu H, Houser SH, Ding J, Chen H, Zhang X, et al. Exploring association between certified EHRs adoption and patient experience in U.S. psychiatric hospitals. PLoS One 2020;15(6):e0234607 [FREE Full text] [doi: 10.1371/journal.pone.0234607] [Medline: 32555623]
91. Gleeson M, Hannigan A, Jamali R, Su Lin K, Klimas J, Mannix M, et al. Using electronic medical records to determine prevalence and treatment of mental disorders in primary care: a database study. Ir J Psychol Med 2016;33(1):3-12. [doi: 10.1017/ipm.2015.10] [Medline: 30115173]
92. Huerta ST. What key factors of electronic health records (EHR) impact mental health professional? satisfaction: a 6 month post-implementation review in a mental health facility. Bradman University. 2015. URL: https://www.proquest.com/openview/b2209cdd89c1dc31d0a592682ef48f30/1.pdf?pq-origsite=gscholar&cbl=18750 [accessed 2018-10-01]

XSL•FO
RenderX

93. Jones AL, Pettey WB, Carter ME, Brignone E, Redd A, Suo Y, et al. Regional variations in documentation of sexual trauma concepts in electronic medical records in the United States veterans health administration. AMIA Annu Symp Proc 2020;2019:514-522 [FREE Full text] [Medline: 32308845]
94. Larrison CR, Xiang X, Gustafson M, Lardiere MR, Jordan N. Implementation of electronic health records among community mental health agencies. J Behav Health Serv Res 2018;45(1):133-142. [doi: 10.1007/s11414-017-9556-9] [Medline: 28439789]
95. Madden JM, Lakoma MD, Rusinak D, Lu CY, Soumerai SB. Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Inform Assoc 2016;23(6):1143-1149 [FREE Full text] [doi: 10.1093/jamia/ocw021] [Medline: 27079506]
96. Wilk JE, Herrell RK, Carr AL, West JC, Wise J, Hoge CW. Diagnosis of PTSD by army behavioral health clinicians: are diagnoses recorded in electronic health records? Psychiatr Serv 2016;67(8):878-882. [doi: 10.1176/appi.ps.201500292] [Medline: 26975516]
97. Wu LT, Payne EH, Roseman K, Case A, Nelson C, Lindblad R. Using a health information technology survey to explore the availability of addiction treatment data in the electronic health records: a National Drug Abuse Treatment Clinical Trials Network study. J Subst Abuse Treat 2020;112S:56-62 [FREE Full text] [doi: 10.1016/j.jsat.2020.01.015] [Medline: 32220412]
98. Brown LA, Mu W, McCann J, Durborow S, Blank MB. Under-documentation of psychiatric diagnoses among persons living with HIV in electronic medical records. AIDS Care 2021;33(3):311-315 [FREE Full text] [doi: 10.1080/09540121.2020.1713974] [Medline: 31931621]
99. Takian A, Sheikh A, Barber N. We are bitter, but we are better off: case study of the implementation of an electronic health record system into a mental health hospital in England. BMC Health Serv Res 2012;12:484 [FREE Full text] [doi: 10.1186/1472-6963-12-484] [Medline: 23272770]
100. 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] [Medline: 24454678]
101. Matthews EB. Integrating the electronic health record into behavioral health encounters: strategies, barriers, and implications for practice. Adm Policy Ment Health 2017;44(4):512-523. [doi: 10.1007/s10488-015-0676-3] [Medline: 26208693]
102. Jung SY, Hwang H, Lee K, Lee D, Yoo S, Lim K, et al. User perspectives on barriers and facilitators to the implementation of electronic health records in behavioral hospitals: qualitative study. JMIR Form Res 2021;5(4):e18764 [FREE Full text] [doi: 10.2196/18764] [Medline: 33830061]
103. Boyer L, Samuelian JC, Fieschi M, Lancon C. Implementing electronic medical records in a psychiatric hospital: a qualitative study. Int J Psychiatry Clin Pract 2010;14(3):223-227. [doi: 10.3109/13651501003717243] [Medline: 24917324]
104. Gibson B, Butler J, Zirkle M, Hammond K, Weir C. Foraging for information in the EHR: the search for adherence related information by mental health clinicians. AMIA Annu Symp Proc 2017;2016:600-608 [FREE Full text] [Medline: 28269856]
105. Hardstone G, Hartswood M, Procter R, Slack R, Voss A, Rees G. Supporting informality: team working and integrated care records. In: Proceedings of the 2004 ACM Conference on Computer supported Cooperative Work. 2004 Presented at: CSCW '04; November 6-10, 2004; Chicago, IL p. 142-151. [doi: 10.1145/1031607.1031632]
106. Zhou X, Ackerman MS, Zheng K. Doctors and psychosocial information: records and reuse in inpatient care. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10; April 10-15, 2010; Atlanta, GA p. 1767-1776 URL: https://doi.org/10.1145/1753326.1753592 [doi: 10.1145/1753326.1753592]
107. Robertson A, Cresswell K, Takian A, Petrakaki D, Crowe S, Cornford T, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. BMJ 2010;341:c4564 [FREE Full text] [doi: 10.1136/bmj.c4564] [Medline: 20813822]
108. Sheikh A, Cornford T, Barber N, Avery A, Takian A, Lichtner V, et al. Implementation and adoption of nationwide electronic health records in secondary care in England: final qualitative results from prospective national evaluation in "early adopter" hospitals. BMJ 2011;343:d6054 [FREE Full text] [doi: 10.1136/bmj.d6054] [Medline: 22006942]
109. Cofiel L, Ray RK, Dias R, Roedel PM, Gutierrez MA, Brentani H. Electronic health records needs from the psychiatry point of view: the result of a 18 month implementation experience. In: Proceedings of the IADIS International Conference e-Health. 2013 Presented at: MCCSIS '13; July 24-26, 2013; Prague, Czech Republic p. 177-179 URL: https://observatorio.fm.usp.br/handle/OPI/18521
110. Jetelina KK, Woodson TT, Gunn R, Muller B, Clark KD, DeVoe JE, et al. Evaluation of an electronic health record (EHR) tool for integrated behavioral health in primary care. J Am Board Fam Med 2018;31(5):712-723 [FREE Full text] [doi: 10.3122/jabfm.2018.05.180041] [Medline: 30201667]
111. Stanhope V, Matthews EB. Delivering person-centered care with an electronic health record. BMC Med Inform Decis Mak 2019;19(1):168 [FREE Full text] [doi: 10.1186/s12911-019-0897-6] [Medline: 31438960]
112. Cifuentes M, Davis M, Fernald D, Gunn R, Dickinson P, Cohen DJ. Electronic health record challenges, workarounds, and solutions observed in practices integrating behavioral health and primary care. J Am Board Fam Med 2015;28 Suppl 1(Suppl 1):S63-S72 [FREE Full text] [doi: 10.3122/jabfm.2015.S1.150133] [Medline: 26359473]

XSL•FO
RenderX

113. Reitz R, Common K, Fifield P, Stiasny E. Collaboration in the presence of an electronic health record. Fam Syst Health 2012;30(1):72-80. [doi: 10.1037/a0027016] [Medline: 22429079]
114. Martin K, Ham E, Hilton NZ. Documentation of psychotropic pro re nata medication administration: an evaluation of electronic health records compared with paper charts and verbal reports. J Clin Nurs 2018;27(15-16):3171-3178. [doi: 10.1111/jocn.14511] [Medline: 29752835]
115. Lipford K, Jones S, Johnson K. Needs assessment of an electronic health record at an inpatient psychiatric hospital. Online J Nurs Inform 2017;21(1):- [FREE Full text] [doi: 10.1007/1-84628-115-6_2]
116. Cellucci LW, Cellucci T, Stanton M, Kerrigan D, Madrake M. Current status and future directions of EMR use in psychology clinics. Health Policy Technol 2015;4(2):91-99. [doi: 10.1016/j.hlpt.2015.02.005]
117. Bell JR, Kilic C, Prabakaran R, Wang YY, Wilson R, Broadbent M, et al. Use of electronic health records in identifying drug and alcohol misuse among psychiatric in-patients. Psychiatrist 2013;37(1):15-20 [FREE Full text] [doi: 10.1192/pb.bp.111.038240]
118. Talmon J, Ammenwerth E, Brender J, De Keizer N, Nykänen P, & Rigby M. STARE-HI - Statement on reporting of evaluation studies in Health Informatics. International journal of medical informatics 2009;78(1):1-9 [FREE Full text]
119. Hay P, Wilton K, Barker J, Mortley J, Cumerlato M. The importance of clinical documentation improvement for Australian hospitals. Health Inf Manag 2020;49(1):69-73. [doi: 10.1177/1833358319854185] [Medline: 31303061]
120. Kuhn T, Basch P, Barr M, Yackel T, Medical Informatics Committee of the American College of Physicians. Clinical documentation in the 21st century: executive summary of a policy position paper from the American College of Physicians. Ann Intern Med 2015;162(4):301-303 [FREE Full text] [doi: 10.7326/M14-2128] [Medline: 25581028]
121. Mamykina L, Vawdrey DK, Stetson PD, Zheng K, Hripcsak G. Clinical documentation: composition or synthesis? J Am Med Inform Assoc 2012;19(6):1025-1031 [FREE Full text] [doi: 10.1136/amiajnl-2012-000901] [Medline: 22813762]
122. Alvarez RC, Zelmer J. Standardization in health informatics in Canada. Int J Med Inform 1998;48(1-3):13-18. [doi: 10.1016/s1386-5056(97)00103-2] [Medline: 9600397]
123. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2019;40:487-500. [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]
124. Browne G. Outcome measures: do they fit with a recovery model? Int J Ment Health Nurs 2006;15(2):153-154. [doi: 10.1111/j.1447-0349.2006.00417.x] [Medline: 16643351]
125. Varpio L, Rashotte J, Day K, King J, Kuziemsky C, Parush A. The EHR and building the patient's story: a qualitative investigation of how EHR use obstructs a vital clinical activity. Int J Med Inform 2015;84(12):1019-1028. [doi: 10.1016/j.ijmedinf.2015.09.004] [Medline: 26432683]
126. Rigby M, Lindmark J, Furlan PM. The importance of developing an informatics framework for mental health. Health Policy 1998;45(1):57-67. [doi: 10.1016/s0168-8510(98)00028-1] [Medline: 10183013]
127. Johnson SB, Bakken S, Dine D, Hyun S, Mendonça E, Morrison F, et al. An electronic health record based on structured narrative. J Am Med Inform Assoc 2008;15(1):54-64 [FREE Full text] [doi: 10.1197/jamia.M2131] [Medline: 17947628]
128. Ho YX, Gadd CS, Kohorst KL, Rosenbloom ST. A qualitative analysis evaluating the purposes and practices of clinical documentation. Appl Clin Inform 2014;5(1):153-168 [FREE Full text] [doi: 10.4338/ACI-2013-10-RA-0081] [Medline: 24734130]
129. Stanhope V, Ingoglia C, Schmelter B, Marcus SC. Impact of person-centered planning and collaborative documentation on treatment adherence. Psychiatr Serv 2013;64(1):76-79. [doi: 10.1176/appi.ps.201100489] [Medline: 23280459]
130. Colicchio TK, Dissanayake PI, Cimino JJ. The anatomy of clinical documentation: an assessment and classification of narrative note sections format and content. AMIA Annu Symp Proc 2021;2020:319-328 [FREE Full text] [Medline: 33936404]
131. Recommendations regarding sensitive health information. National Committee of Vital and Health Statistics. 2010. URL: https://www.ncvhs.hhs.gov/wp-content/uploads/2014/05/101110lt.pdf [accessed 2021-03-10]
132. Cairns I, Jonas M, Wallis K. The ethics of sharing: how do social workers decide what to record in shared health records? Ethics Soc Welf 2018;12(4):348-369. [doi: 10.1080/17496535.2017.1384849]
133. Langan J. Mental health, risk communication and data quality in the electronic age. Br J Soc Work 2007;39(3):467-487. [doi: 10.1093/bjsw/bcm129]
134. Drake RE, Deegan PE, Rapp C. The promise of shared decision making in mental health. Psychiatr Rehabil J 2010;34(1):7-13. [doi: 10.2975/34.1.2010.7.13] [Medline: 20615839]
135. Pisciotta M, Denneson LM, Williams HB, Woods S, Tuepker A, Dobscha SK. Providing mental health care in the context of online mental health notes: advice from patients and mental health clinicians. J Ment Health 2019;28(1):64-70. [doi: 10.1080/09638237.2018.1521924] [Medline: 30468100]
136. Maniss S, Pruit AG. Collaborative documentation for behavioral healthcare providers: an emerging practice. J Hum Services Training Res Pract 2018;3(1):2 [FREE Full text]
137. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. Int J Med Inform 2006;75(7):530-541. [doi: 10.1016/j.ijmedinf.2005.08.009] [Medline: 16198142]

XSL•FO
RenderX

138. Crampton NH, Reis S, Shachak A. Computers in the clinical encounter: a scoping review and thematic analysis. J Am Med Inform Assoc 2016;23(3):654-665 [FREE Full text] [doi: 10.1093/jamia/ocv178] [Medline: 26769911]
139. De Lusignan S, Kumarapeli P, Debar S, Kushniruk AW, Pearce C. Using an open source observational tool to measure the influence of the doctor's consulting style and the computer system on the outcomes of the clinical consultation. Stud Health Technol Inform 2009;150:1017-1021. [Medline: 19745467]
140. Shachak A, Reis S. The impact of electronic medical records on patient-doctor communication during consultation: a narrative literature review. J Eval Clin Pract 2009;15(4):641-649. [doi: 10.1111/j.1365-2753.2008.01065.x] [Medline: 19522722]
141. Pearce C, Arnold M, Phillips C, Trumble S, Dwan K. The patient and the computer in the primary care consultation. J Am Med Inform Assoc 2011;18(2):138-142 [FREE Full text] [doi: 10.1136/jamia.2010.006486] [Medline: 21262923]
142. McGinn CA, Grenier S, Duplantie J, Shaw N, Sicotte C, Mathieu L, et al. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Med 2011;9:46 [FREE Full text] [doi: 10.1186/1741-7015-9-46] [Medline: 21524315]
143. Gephart S, Carrington JM, Finley B. A systematic review of nurses' experiences with unintended consequences when using the electronic health record. Nurs Adm Q 2015;39(4):345-356. [doi: 10.1097/NAQ.0000000000000119] [Medline: 26340247]
144. Strudwick G, Eyasu T. Electronic health record use by nurses in mental health settings: a literature review. Arch Psychiatr Nurs 2015;29(4):238-241. [doi: 10.1016/j.apnu.2015.03.007] [Medline: 26165979]
145. Roman LC, Ancker JS, Johnson SB, Senathirajah Y. Navigation in the electronic health record: a review of the safety and usability literature. J Biomed Inform 2017;67:69-79 [FREE Full text] [doi: 10.1016/j.jbi.2017.01.005] [Medline: 28088527]
146. Ratwani R, Fairbanks T, Savage E, Adams K, Wittie M, Boone E, et al. Mind the gap. A systematic review to identify usability and safety challenges and practices during electronic health record implementation. Appl Clin Inform 2016;7(4):1069-1087 [FREE Full text] [doi: 10.4338/ACI-2016-06-R-0105] [Medline: 27847961]
147. Kruse CS, Mileski M, Alaytsev V, Carol E, Williams A. Adoption factors associated with electronic health record among long-term care facilities: a systematic review. BMJ Open 2015;5(1):e006615 [FREE Full text] [doi: 10.1136/bmjopen-2014-006615] [Medline: 25631311]
148. Spiranovic C, Matthews A, Scanlan J, Kirkby KC. Increasing knowledge of mental illness through secondary research of electronic health records: opportunities and challenges. Adv Ment Health 2016;14(1):14-25. [doi: 10.1080/18387357.2015.1063635]
149. Smoller JW. The use of electronic health records for psychiatric phenotyping and genomics. Am J Med Genet B Neuropsychiatr Genet 2018;177(7):601-612 [FREE Full text] [doi: 10.1002/ajmg.b.32548] [Medline: 28557243]
150. Barak-Corren Y, Castro VM, Javitt S, Hoffnagle AG, Dai Y, Perlis RH, et al. Predicting suicidal behavior from longitudinal electronic health records. Am J Psychiatry 2017;174(2):154-162. [doi: 10.1176/appi.ajp.2016.16010077] [Medline: 27609239]
151. Gooding P. Mapping the rise of digital mental health technologies: emerging issues for law and society. Int J Law Psychiatry 2019;67:101498. [doi: 10.1016/j.ijlp.2019.101498] [Medline: 31785726]
152. Gooding P, Kariotis T. Ethics and law in research on algorithmic and data-driven technology in mental health care: scoping review. JMIR Ment Health 2021;8(6):e24668 [FREE Full text] [doi: 10.2196/24668] [Medline: 34110297]
153. Zhang J, Walji MF. TURF: toward a unified framework of EHR usability. J Biomed Inform 2011;44(6):1056-1067 [FREE Full text] [doi: 10.1016/j.jbi.2011.08.005] [Medline: 21867774]
154. O Brien A, Weaver C, Settergren TT, Hook ML, Ivory CH. EHR documentation: the hype and the hope for improving nursing satisfaction and quality outcomes. Nurs Adm Q 2015;39(4):333-339. [doi: 10.1097/NAQ.0000000000000132] [Medline: 26340245]
155. 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] [Medline: 31735343]
156. Domaney NM, Torous J, Greenberg WE. Exploring the association between electronic health record use and burnout among psychiatry residents and faculty: a pilot survey study. Acad Psychiatry 2018;42(5):648-652. [doi: 10.1007/s40596-018-0939-x] [Medline: 29785625]
157. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
158. Walker L, Clendon J. The case for end-user involvement in design of health technologies. J Telemed Telecare 2016;22(8):443-446. [doi: 10.1177/1357633X16670479] [Medline: 27799445]
159. Final Report - volume 5: transforming the system—innovation and implementation. Royal Commission into Victoria's Mental Health System. 2021. URL: https://finalreport.rcvmhs.vic.gov.au/download-report/ [accessed 2021-03-01]
160. Electronic medical record adoption model (EMRAM). Healthcare Information and Management Systems Society. URL: https://www.himss.org/what-we-do-solutions/digital-health-transformation/maturity-models/electronic-medical-record-adoption-model-emram [accessed 2019-02-10]

XSL•FO
RenderX

161. Blobel B. Interoperable EHR systems - challenges, standards and solutions. Eur J Biomed Inform 2018;14(2):10-19 [FREE Full text] [doi: 10.24105/ejbi.2018.14.2.3]
162. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform 2019;94:103188 [FREE Full text] [doi: 10.1016/j.jbi.2019.103188] [Medline: 31063828]
163. Bodenreider O, Cornet R, Vreeman DJ. Recent developments in clinical terminologies - SNOMED CT, LOINC, and RxNorm. Yearb Med Inform 2018;27(1):129-139 [FREE Full text] [doi: 10.1055/s-0038-1667077] [Medline: 30157516]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
NHS: National Health Service
PRISMA-ScR: Preferred Reporting Item for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by A Mavragani; submitted 13.05.21; peer-reviewed by B Gibson, C Weir, S Sadasivaiah; comments to author 28.06.21; revised version received 14.09.21; accepted 13.01.22; published 04.05.22.

Please cite as:
Kariotis TC, Prictor M, Chang S, Gray K
Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review
J Med Internet Res 2022;24(5):e30405
URL: https://www.jmir.org/2022/5/e30405
doi:10.2196/30405
PMID:35507393

©Timothy Charles Kariotis, Megan Prictor, Shanton Chang, Kathleen Gray. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2022. 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.

XSL•FO
RenderX

Review

eHealth Tools That Assess and Track Health and Well-being in Children and Young People: Systematic Review

Elizabeth Stewart[1], BA, PhD, DPsych; Alyssa Milton[1]; Hannah Frances Yee[1]; Michael Jae Song[2]; Anna Roberts[1]; Tracey Davenport[1]; Ian Hickie[1]

[1]Brain and Mind Centre, The University of Sydney, Sydney, Australia

[2]Department of Psychiatry, The University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Elizabeth Stewart, BA, PhD, DPsych
Brain and Mind Centre, The University of Sydney
94-100 Mallett Street Camperdown
Sydney, 2050
Australia
Phone: 61 0404060549
Email: lizziestewart@gmail.com

Abstract

Background: eHealth tools that assess and track health outcomes in children or young people are an emerging type of technology that has the potential to reform health service delivery and facilitate integrated, interdisciplinary care.

Objective: The aim of this review is to summarize eHealth tools that have assessed and tracked health in children or young people to provide greater clarity around the populations and settings in which they have been used, characteristics of digital devices (eg, health domains, respondents, presence of tracking, and connection to care), primary outcomes, and risks and challenges of implementation.

Methods: A search was conducted in PsycINFO, PubMed or MEDLINE, and Embase in April 2020. Studies were included if they evaluated a digital device whose primary purpose was to assess and track health, focused on children or young people (birth to the age of 24 years), reported original research, and were published in peer-reviewed journals in English.

Results: A total of 39 papers were included in this review. The sample sizes ranged from 7 to 149,329 participants (median 163, mean 5155). More studies were conducted in urban (18/39, 46%) regions than in rural (3/39, 8%) regions or a combination of urban and rural areas (8/39, 21%). Devices were implemented in three main settings: outpatient health clinics (12/39, 31%), hospitals (14/39, 36%), community outreach (10/39, 26%), or a combination of these settings (3/39, 8%). Mental and general health were the most common health domains assessed, with a single study assessing multiple health domains. Just under half of the devices tracked children's health over time (16/39, 41%), and two-thirds (25/39, 64%) connected children or young people to clinical care. It was more common for information to be collected from a single informant (ie, the child or young person, trained health worker, clinician, and parent or caregiver) than from multiple informants. The health of children or young people was assessed as a primary or secondary outcome in 36% (14/39) of studies; however, only 3% (1/39) of studies assessed whether using the digital tool improved the health of users. Most papers reported early phase research (formative or process evaluations), with fewer outcome evaluations and only 3 randomized controlled trials. Identified challenges or risks were related to accessibility, clinical utility and safety, uptake, data quality, user interface or design aspects of the device, language proficiency or literacy, sociocultural barriers, and privacy or confidentiality concerns; ways to address these barriers were not thoroughly explored.

Conclusions: eHealth tools that assess and track health in children or young people have the potential to enhance health service delivery; however, a strong evidence base validating the clinical utility, efficacy, and safety of tools is lacking, and more thorough investigation is needed to address the risks and challenges of using these emerging technologies in clinical care. At present, there is greater potential for the tools to facilitate multi-informant, multidomain assessments and longitudinally track health over time and room for further implementation in rural or remote regions and community settings around the world.

(J Med Internet Res 2022;24(5):e26015) doi:10.2196/26015

KEYWORDS

eHealth; children; young people; health; technology; mobile phone

XSL•FO
RenderX

Introduction

In 2018, the United Nations Children's Fund released a report on digital technologies in health [1] alongside its Strategic Plan, 2018-2021, which emphasized the importance of using digital (internet- and mobile-based) technology to facilitate health care for children and young people. At the same time, the Early Childhood Developmental Interventions Review Group for the Lancet issued a report making several recommendations for the improvement of assessments and interventions for children or young people [2]. These recommendations included improving the capacity for services to conduct multidomain and multi-informant assessments, connecting children or young people and families with personalized care options, and using digital solutions within health services to allow for broader-scale change [2]. Together, these reports highlight the emerging need to use digital technologies to enhance the delivery of health care for children or young people and their families.

Over the past decade, there has been a rapid growth in the development of digital tools in the health and well-being space [3]. These tools have served various purposes in health care, with the most common uses among children or young people being to deliver interventions (eg, clinician-assisted evidence-based treatments and self-monitoring and self-care), provide education, and facilitate communication for both consumers and clinicians (eg, telehealth or teleconferencing and online peer support groups) [2,4]. Another more recent use of eHealth has been to facilitate the assessment and triage of children or young people through health services [5-12]. These emerging technologies provide users (ie, clinicians and consumers) with secure, web-based platforms for submitting health data (sometimes automatically via biosensors or wearables) without having to be physically present in a hospital or health clinic. The information can be securely shared with health professionals with expertise in children or young people's areas of need, allowing them to be triaged to appropriate services and connected with ongoing care [5,7-10,13,14]. Thus, these eHealth solutions differ from existing technologies in that their goal is not to deliver interventions or ongoing treatment but rather to facilitate a connection between consumers and pre-existing health services, allow for routine outcome monitoring, and place the person (or family) at the center of care. Furthermore, although some of these tools provide education resources (eg, fact sheets) or communication pathways (eg, web-based chats), they do this with the goal of triaging children or young people to appropriate care.

The literature on eHealth tools that assess and track health outcomes in children or young people is still in its infancy; however, a growing number of studies have reported on such devices over the past decade [4,15]. These tools differ in their health focus (eg, infectious diseases and mental health) [16,17] and locations in which they have been used (ie, rural or urban areas, high- or low-income countries, and specific health settings) [7,18-20]. There has also been variability in terms of the respondent who enters data into the tool (ie, clinician and consumer), the type of data (ie, questionnaires and physiological data), whether the tools have facilitated only assessment or assessment and tracking, and whether they have connected children or young people to clinical care. Given this variability, the specific features of eHealth tools, as well as their efficacy for improving health outcomes and clinical care delivery for children or young people, remain unclear.

Despite the potential benefits of eHealth solutions for children or young people, numerous challenges have been documented in their development, implementation, and uptake among other groups [10,13,21-24]. To be successful, the technologies must be user-friendly, engaging, and accessible to diverse populations. Issues of language, literacy, and culture have all been found to affect accessibility, uptake, and the quality of data [12,21,22,25]. The validity and integrity of data also depend on the availability of appropriately trained health care workers to enter or interpret information, emphasizing the importance of developing and evaluating these tools within the contexts in which they will be used. Finally, issues of privacy, confidentiality, and security are paramount to ensuring that the tools respect the rights of users and are likely to affect the uptake of these technologies [22,25,26].

To our knowledge, no comprehensive reviews have been conducted to examine the efficacy of eHealth tools that assess and track health outcomes in children or young people. As such, it is unclear in which health domains and settings these tools may have the potential to shape clinical care and, importantly, whether their use has been associated with improved health outcomes for children or young people. There is also a need to identify potential challenges and risks of using eHealth tools to ensure that best practice methods are established and consistently used [1,27]. Understanding the available eHealth solutions and their efficacies is critical for shaping future research and development efforts and ensuring efficient expansion of knowledge in this field.

The aim of this systematic review is to summarize eHealth tools that have been developed to assess and track health in children or young people to provide greater clarity about (1) the populations and settings in which these tools have been used (ie, locations, languages, and age groups); (2) characteristics of the tools (ie, health domains assessed, respondent, type of data, presence of tracking, and connection to care); (3) primary outcomes of the study, including whether the use of the tool has been associated with improved health outcomes; and (4) risks and challenges identified during implementation and evaluation.

Methods

Overview

The term eHealth has been variously defined in the literature. According to a systematic review, 51 unique definitions have been used for the term, without a clear consensus on a single definition, and the definitions differ in how inclusively they are conceptualized [28,29]. We have chosen to use a definition based on the conceptualization of eHealth offered by Vegesna et al [30] because of its relevance and consistency with the overarching aims of this review; digital technologies are thus defined as noninvasive digital devices that have been used to assess and track the health of a patient or consumer. We used the World Health Organization's definition of childhood and

XSL•FO
RenderX

youth as the period spanning birth to 24 years, whereby children are aged 0 to 9 years, and young people are aged 10 to 24 years [31].

Search Strategy

The search was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [32]. PsycINFO, PubMed or MEDLINE, and Embase were searched via OVID by 3 members of the research team (ES, HY, and AR) on April 27, 2020. The following terms were used *((child*)* OR *(adolescen*)* OR *(young person)* OR *(infan*))* AND *((wellbeing)* OR *(health))* AND *(((digital tool)* OR *(digital* AND *tool))* OR *(eHealth)* OR *((mobile application)* OR *(mobile* AND *application)))*. A wildcard (*) was placed at the end of each applicable search term to ensure that all relevant terms were captured. All Medical Subject Heading terms were explored to broaden the search for relevant studies. Date limits were not set on any of the database searches. The reference lists of relevant reviews and identified empirical studies were searched to identify further studies, as per the ancestry method.

Study Selection Criteria

Studies were included if they met the following criteria:

1. Included children and young people (birth to the age of 24 years) or their parents or carers, as per the World Health Organization's definition [31]
2. Evaluated a digital device, including internet- or mobile-based technology (ie, noninvasive digital devices, including internet- or mobile-based e-tools and wearable devices), the primary purpose of which was to assess or track the health of the child or young person
3. Focused on a domain of health
4. Evaluation studies, meaning the authors evaluated some aspects of the digital device, including effectiveness, validity, or feasibility; we included all or any type of evaluation studies, which were categorized according to the Center for Disease Control definition (ie, formative, process, and outcome) [33] and National Health and Medical Research Council criteria for study design [34]
5. Reported original research
6. Published in English in a peer-reviewed journal and included human participants

Studies were excluded if they had the following characteristics:

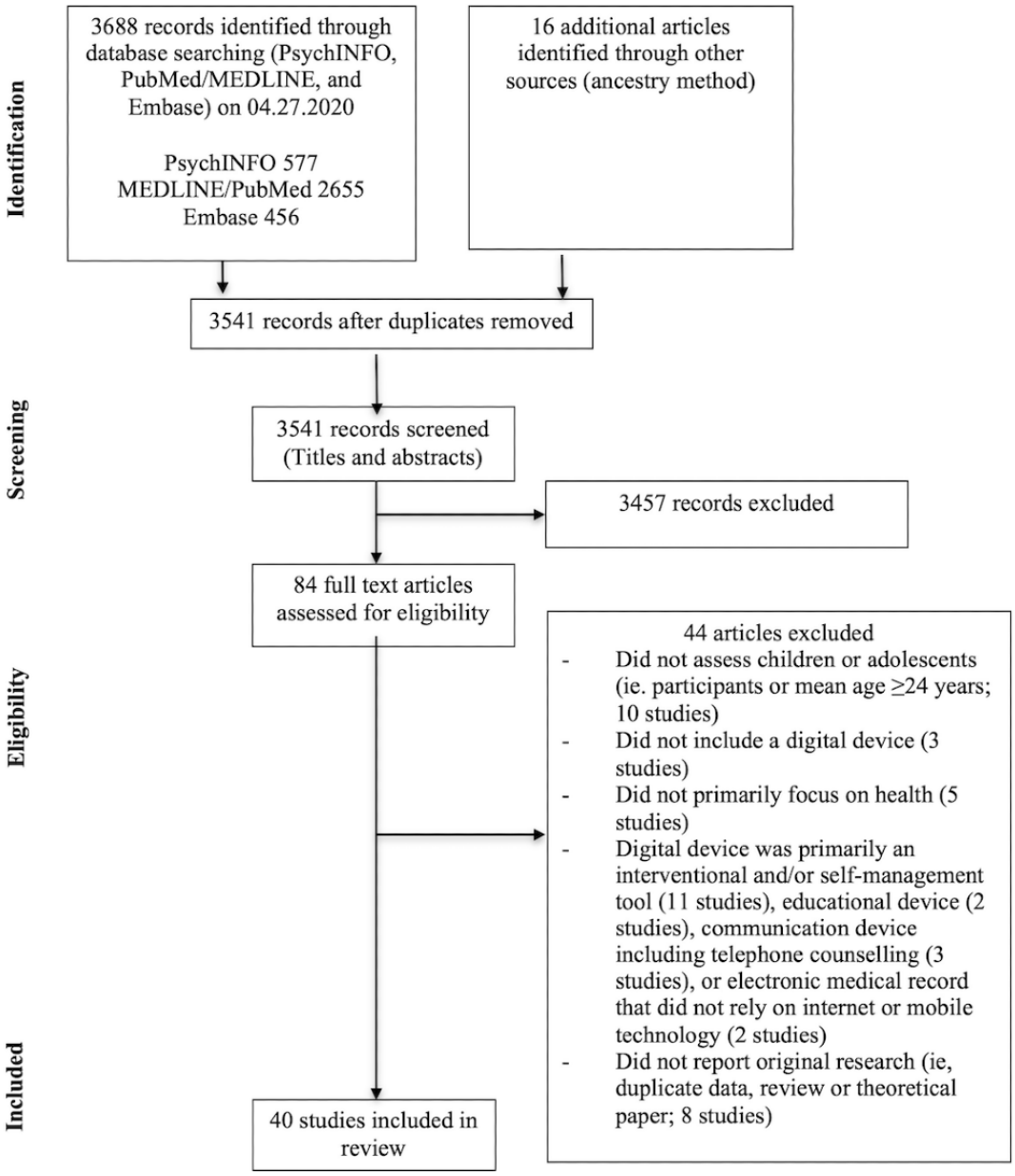
1. Included adults only with no child or young person focus or if >25% of participants were outside our age criteria (birth to the age of 24 years)
2. Evaluated a digital device that was primarily an interventional tool (ie, clinician-led and self-management tools), an educational device (eg, an e-course), a communication device (ie, assistive communication with images or written or spoken language; and teleconferencing only without additional assessment or tracking functionality), or digital technology that did not use internet or mobile technology (eg, electronic medical record systems)
3. Reported results from development or description of the tool that had not yet been evaluated (eg, protocol papers)

Procedure

Figure 1 displays the process of study selection. The search retrieved 3688 papers, and an additional 16 papers were identified by searching the reference lists of relevant papers and reviews. Of these papers, 95.59% (3541/3704) remained after duplicates were removed. The titles and abstracts of all papers were screened by 2 reviewers (ES and HY). Of the 3541 papers, 84 (2.37%) full-text papers met the inclusion criteria and were obtained. A conservative approach was taken to ensure that relevant papers were not missed, and full-text papers were reviewed if the reviewers could not determine with certainty whether the inclusion criteria were met. The manuscripts of these 84 papers were reviewed by 2 independent raters (ES and HY), and discrepancies were resolved via discussion. Of the 84 papers, 45 (54%) papers were excluded, leaving 39 (46%) papers that were included in the review.

XSL•FO
RenderX

Figure 1. Flow diagram of identification and selection of studies.



Data Extraction

The following information was extracted from each paper:

1. Name of the first author, year of publication, and age range of children or young people
2. Language or languages used in the digital application
3. Location where the study was conducted: country, locality (urban or rural), and setting; locality was defined according to the Organization for Economic Co-operation and Development harmonized definition of global urbanization, which uses the population density of the area, that is, rural (<5000 inhabitants) or urban (≥5000 inhabitants) [35]; some studies were conducted in multiple locations, which was considered in categorizing study locality as urban, rural, or a mixture of urban and rural settings
4. Characteristics of the digital tool: health domain assessed, respondent (parent or caregiver, child or young person, clinician, trained health worker, and other), device type (mobile, desktop, and tablet), type of data (questionnaire or survey, images, and physiological), whether the tool allowed for tracking over time (ie, data collected at multiple time points), and whether the device facilitated connection to care (ie, linking patients to health care providers or services)
5. Study characteristics: type of evaluation study, defined according to the Center for Disease Control definition of study evaluation types, that is, formative, process, or evaluation [33]; study type: qualitative, quantitative, or mixed methods; and (3) study design, based on the National Health and Medical Research Council guidelines [34]
6. The primary outcome and main findings from the study, including whether the health of the child or young person was measured as an outcome in the study
7. Funding source, categorized as public sector (ie, government, universities, research institutes, and professional associations), commercial or not-for-profit (NFP) organizations; these categories were guided by an

XSL•FO
RenderX

Australian Government resource on university research funding (REF)

8. Any documented risks or challenges associated with the use of the eHealth tool

Data Analysis

Descriptive analyses were used to summarize variables of interest, including health domain, location, language, type of data, intended user, presence of certain features (ie, tracking over time and connection to care), and type of evaluation. Frequency data and percentages were used to examine and compare studies on key outcome measures. This approach to analysis was taken because of considerable variability in study objectives and designs and as most studies reported simple quantitative, descriptive statistics or qualitative findings.

Quality Appraisal of Studies

To evaluate the methodological quality of the studies, 2 checklists were used. The Downs and Black checklist [36] was completed for quantitative studies, which measures the quality of both randomized and nonrandomized studies evaluating novel health interventions. The National Institute for Health and Care Excellence Quality Appraisal Checklist was completed for studies reporting qualitative findings [37]. Studies reporting both qualitative and quantitative data were appraised using both checklists. A full description of the checklists and scoring criteria is included in Multimedia Appendix 1 [5-9,16-20,38-66].

Results

Demographics of Studies

Table 1 summarizes the characteristics of the 39 studies included in this review. All (38/39, 97%) but a single (1/39, 3%) study was published in the past decade (2010-2020), and over one-third of the studies (15/39, 39%) were published in the past year (2019-2020; Figure 2). Most studies were conducted in a single country (35/39, 90%), most commonly America (7/39, 18%) or Australia (6/39, 15%). English was the sole language of communication in 49% (19/39) of studies; 13% (5/39) of studies evaluated tools that used English and at least one other language, and 21% (8/39) used languages other than English; the remaining 18% (7/39) of studies did not report enough information to determine which language was used in the tool. Regarding locality, studies were conducted in urban (18/39, 46%), rural (3/39, 8%), or a mixture of urban and rural settings (8/39, 21%); 26% (10/39) studies did not report enough information to determine locality. Digital devices were implemented across 3 main settings: outpatient health clinics (12/39, 31%), hospitals (ie, inpatient units and emergency departments; 14/39, 36%), and community outreach (ie, community spaces that were not formal health clinics; 10/39, 26%) or a combination of these settings (3/39, 8%).

XSL•FO
RenderX

Table 1. Demographic characteristics of studies.

Study	Age range of children	Country	Locality[a]	Language used in the device	Setting
Alawna et al, 2019 [50]	19-27 years (mean 22.0)	Turkey	NR[b]	NR	Outpatient health clinic
Binotti et al, 2019 [56]	Infants (age range NR)	Italy	Urban	NR	Hospital
Boyce et al, 2019 [48]	2-59 months	Malawi	NR	English	Outpatient health clinic
Den Boer et al, 2018 [18]	3-17 years	Netherlands	Urban (81%) and rural (19%)	English, German, Spanish, and Dutch	Outpatient health clinic
Detsomboonrat and Pisarntu-rakit, 2019 [58]	Children in primary school (age range NR)	Thailand	NR	Thai	Community outreach
Dexheimer et al, 2014 [19]	2-18 years	United States	Urban	English	Hospital
Eikelboom et al, 2005 [5]	9 months-16 years	Australia	Rural	English	Outpatient health clinic
Estai et al, 2016 [57]	2-18 years	Australia	Urban	English	Outpatient health clinic
Finocchario-Kessler et al, 2015 [6]	Children (age range NR)	Kenya	Urban (50%) and rural (50%)	English	Hospital
Franke et al, 2018 [20]	18 months-14 years	Ghana	Urban	Twi	Hospital
Galvez et al 2017 [52]	Children (age range NR)	55 countries (world-wide)	NR	English	Hospital
Ginsburg et al, 2015 [16]	Children (age range NR)	Ghana	Urban	English	Outpatient health clinic
Gregory et al, 2017 [39]	<18 years	United Kingdom	Urban	English	Hospital
Han et al, 2019 [53]	13-26 years	China and Australia	Urban	NR	Hospital
Hashemi et al, 2017 [7]	6-18 years	Gaza	Urban	English and Arabic	Community outreach
Heida et al, 2018 [62]	10-19 years	Netherlands	Urban (55%) and rural (45%)	Dutch	Outpatient health clinic
Hussey and Flynn, 2019 [41]	0-21 years	United States	Urban	English	Outpatient health clinic
Iorfino et al, 2017 [8]	16-24 years	Australia	Urban (85%) and rural (15%)	English	Outpatient health clinic
Jeong et al, 2020 [40]	15-19 years	South Korea	Urban	Korean	Hospital, outpatient health clinic, and community out-reach
Jiam et al, 2017 [66]	3-22 years	United States	NR	English	Community outreach
Kassam-Adams et al, 2019 [42]	6-14 years	United States	Urban (50%) and rural (50%)	English	Hospital
Kim et al, 2019 [60]	0-5 years	South Korea	NR	NR	Community outreach
Li et al, 2019 [63]	1-18 years	China	Urban	Mandarin	Hospital
March et al, 2018 [17]	5-12 years	Australia	Urban	English	Outpatient health clinic and community outreach
Matin et al, 2020 [59]	0-7 days	Uganda	Rural	Lusoga and English	Community outreach
McCulloh et al, 2018 [49]	0-2 months	United States	NR	English	Hospital
Mohammed et al, 2018 [9]	0-5 years	Ghana	Urban	Twi	Community outreach
Padidar et al, 2019 [64]	0-9 days	Iran	Urban	NR	Hospital
Rath et al, 2018 [45]	0-24 years	Germany	NR	Arab, Farsi, and Russian	Community outreach
Rath et al, 2019 [65]	0-5 years	Germany and Greece	Urban	NR	Hospital

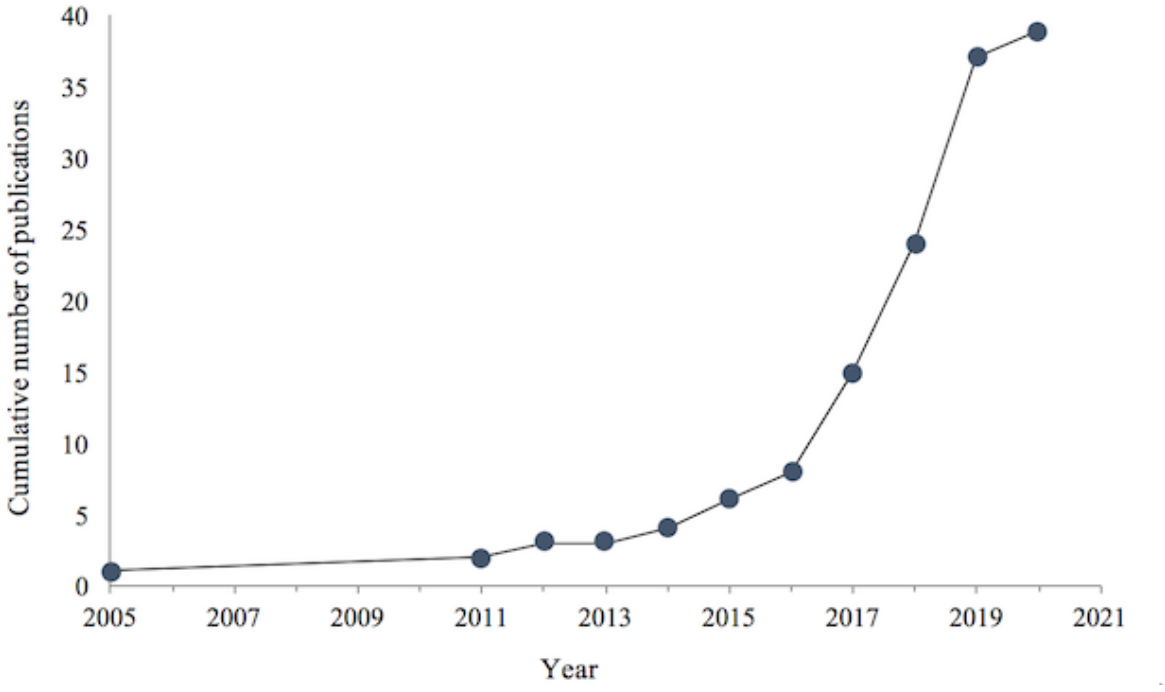
XSL•FO
RenderX

Study	Age range of children	Country	Locality[a]	Language used in the device	Setting
Reid et al, 2011 [43]	14-24 years	Australia	Urban (50%) and rural (50%)	English	Outpatient health clinic
Singh et al, 2017 [51]	0-2 years	India	Urban (85%) and rural (15%)	Hindi, Gujarati, and English	Hospital and outpatient health clinic
Svedberg et al, 2019 [46]	6-13 years	Sweden	Urban (50%) and rural (50%)	Swedish	Hospital
Thabrew et al, 2019 [44]	13-14 years	New Zealand	NR	English	Outpatient health clinic
Thabtah, 2018 [54]	0-17 years	10 countries	NR	11 languages	Community outreach
Thompson et al, 2016 [47]	12-18 years	United States	Urban	English	Outpatient health clinic
Valdes-Angues et al, 2018 [55]	3-18 years	Uganda and United States	Rural	English	Community outreach
van Karnebeek et al, 2012 [61]	0-18 years	Canada	Urban	English	Hospital
Wang et al, 2017 [38]	5-17 years	China	Urban	English and Chinese	Outpatient health clinic

[a]Locality: region in which the eHealth tool was implemented, defined as rural (<5000 inhabitants) or urban (≥5000 inhabitants), according to the Organization for Economic Co-operation and Development's harmonized definition of global urbanization [35].

[b]NR: not reported.

Figure 2. Cumulative number of studies published each year.



Characteristics of eHealth Tools

Table 2 summarizes the characteristics of the studied digital devices.

XSL•FO
RenderX

Table 2. Device characteristics.

Study	Health domain	Device	Type of data	Respondent	Tracking over time	Connection to care
Alawna et al, 2019 [50]	General health	Mobile	Physiological	Trained health worker[a]	Unclear	Yes
Binotti et al, 2019 [56]	Developmental	Mobile	Physiological	Trained health worker	No	No
Boyce et al, 2019 [48]	General health	Mobile	Questionnaire or survey	Trained health worker	No	Yes
Den Boer et al, 2018 [18]	Oral health	Mobile	Questionnaire or survey	Clinician[b] and child or young person[c]	No	Yes
Detsomboonrat and Pisarnturakit 2019 [58]	Oral	Mobile and desktop	Questionnaire or survey	Clinician	No	Yes
Dexheimer et al, 2014 [19]	General health	Desktop	Questionnaire or survey	Clinician	Yes	Yes
Eikelboom et al, 2005 [5]	Ear, nose, and throat	Desktop	Images	Clinician	No	Yes
Estai et al, 2016 [57]	Oral health	Desktop	Images	Clinician and trained health worker	No	No
Finocchario-Kessler et al, 2015 [6]	Infectious	Desktop	Physiological	Parent or caregiver[d] and trained health worker	Yes	Yes
Franke et al, 2018 [20]	Infectious	Mobile	Questionnaire or survey	Parent or caregiver	No	Yes
Galvez et al, 2017 [52]	Emergency	Mobile	Questionnaire or survey	Clinician	Yes	No
Ginsburg et al, 2015 [16]	Infectious	Mobile	Physiological	Trained health worker	No	Yes
Gregory et al, 2017 [39]	Mental health	Mobile	Questionnaire or survey	Clinician and child or young person	No	Yes
Han et al, 2019 [53]	Vision	Mobile	Physiological	Child or young person	No	No
Hashemi et al, 2017 [7]	Mental health	Desktop and mobile	Questionnaire or survey	Trained health worker	No	No
Heida et al, 2018 [62]	Physical health	Desktop	Questionnaire or survey and physiological	Child or young person and parent or caregiver	Yes	Yes
Hussey and Flynn, 2019 [41]	Mental health	Mobile	Questionnaire or survey	Clinician and child or young person	Yes	Yes
Iorfino et al, 2017 [8]	Mental health	Desktop	Questionnaire or survey	Child or young person	Yes	Yes
Jeong et al, 2020 [40]	Mental health	Mobile	Questionnaire or survey	Clinician and child or young person	No	Yes
Jiam et al, 2017 [66]	Neurological	Desktop	Questionnaire or survey	Parent or caregiver and child or young person	Yes	No
Kassam-Adams et al, 2019 [42]	Mental health	Mobile	Questionnaire or survey	Child or young person	Yes	No
Kim et al, 2019 [60]	Infectious	Mobile	Questionnaire or survey and physiological	Parent or caregiver	Yes	No
Li et al, 2019 [63]	Surgery	Mobile	Questionnaire or survey	Child or young person	No	Yes
March et al, 2018 [17]	Mental health	Desktop, mobile, and tablet	Questionnaire or survey	Clinician, parent or caregiver, education provider, and child or young person	No	No

XSL•FO
RenderX

Study	Health domain	Device	Type of data	Respondent	Tracking over time	Connection to care
Matin et al, 2020 [59]	Developmental	Mobile	Questionnaire or survey and physiological	Parent or caregiver	Yes	Yes
McCulloh et al, 2018 [49]	General health	Mobile	Questionnaire or survey	Clinician	No	Yes
Mohammed et al, 2018 [9]	General health	Mobile	Questionnaire or survey	Parent or caregiver	No	Yes
Padidar et al, 2019 [64]	Developmental	Mobile	Physiological images	Clinician and parent or caregiver	No	No
Rath et al, 2018 [45]	General health	Mobile tablet	Questionnaire or survey	Child or young person and parent or caregiver	No	No
Rath et al, 2019 [65]	Infectious	Mobile	Questionnaire or survey	Child or young person	No	No
Reid et al, 2011 [43]	Mental health	Desktop and mobile	Questionnaire or survey	Child or young person	Yes	Yes
Singh et al, 2017 [51]	General health	Desktop and mobile	Questionnaire or survey and physiological	Clinician and parent or caregiver	Yes	Yes
Svedberg et al, 2019 [46]	General health	Mobile	Questionnaire or survey	Child or young person	Yes	Yes
Thabrew et al, 2019 [44]	Mental health	Mobile and tablet	Questionnaire or survey	Child or young person	No	Yes
Thabtah, 2018 [54]	Developmental	Mobile	Questionnaire or survey	Clinician and parent or caregiver	No	No
Thompson et al, 2016 [47]	General health	Desktop	Questionnaire or survey and physiological	Parent or caregiver and young person	Yes	Yes
Valdes-Angues et al, 2018 [55]	Neurological	Desktop and mobile	Questionnaire or survey	Trained health worker	Yes	Yes
van Karnebeek et al, 2012 [61]	Developmental	Desktop, mobile, and tablet	Questionnaire or survey	Clinician	No	No
Wang et al, 2017 [38]	General health and mental health	Mobile	Questionnaire or survey	Clinician, parent or caregiver, and child or young person	Yes	Yes

[a]Trained health workers are staff without professional training who received specific training in the use of the digital tool and associated health domain.
[b]Clinician is defined as a health professional with qualifications in a particular field of practice (including medical doctors and allied health workers).
[c]Child or young person is the individual for whom the eHealth tool was developed.
[d]Parent or caregiver is the primary carer of the child or young person.

Health Domains

Mental and general health were the most common eHealth domains assessed, with each evaluated in 26% (10/39) of studies. Other health domains assessed included child development (5/39, 13%), infectious diseases (5/39, 13%), oral health (3/39, 8%), neurological illnesses (2/39, 5%), ear nose and throat (1/39, 3%), emergency medicine (1/39, 3%), physical health (1/39, 3%), vision (1/39, 3%), and pediatric surgery (1/39, 3%). A single study assessed multiple health domains (mental and general health) [38]. Given that mental health was more commonly assessed than other health domains, we examined these studies further to determine their aim or purpose and the type of information collected. Of the 23% (9/39) of studies that solely assessed mental health, 33% (3/9) focused on suicide prevention [8,39,40], 22% (2/9) focused on early intervention and prevention of mental illness [7,17], and 44% (4/9) focused on multidimensional assessment or management of mental health symptoms [41-44]. Of the 23% (9/39) of studies that solely assessed general health, 22% (2/9) focused on symptom detection and monitoring [9,45], 22% (2/9) provided a platform for patients to view and monitor their health information [46,47], 33% (3/9) focused on digitalized tracking of clinical decision-making [19,48,49], and 22% (2/9) were primarily for assessment [50,51].

Data Collection: Respondent, Type of Data, and Device

All devices measured the health of a child or young person; however, devices differed in the person who entered the health information (ie, the respondent: child or young person, parent

XSL•FO
RenderX

or caregiver, clinician, and trained health worker). Under half of the devices collected information from multiple respondents (16/39, 41%); other tools collected information solely from a child or young person (8/39, 21%), clinician (6/39, 15%), trained health worker (6/39, 15%), or parent or caregiver (3/39, 8%). Approximately 15% (6/39) of studies collected data in multiple forms (ie, questionnaire or survey, physiological data, or images); otherwise, data were collected solely in the form of questionnaires or surveys (26/39, 67%), physiological data (5/39, 13%), or images (2/39, 5%). Most eHealth tools (31/39, 80%) were configured to collect data on a mobile phone, of which some (9/39, 23%) were also configured to collect data on another device (ie, desktop or tablet).

Device Features: Health Tracking and Connection to Care

Just under half of the devices tracked children's health over time (16/39, 41%), and two-thirds (25/39, 64%) connected children or young people to clinical care, whereas the remainder did not.

Outcome Evaluation: Primary Outcome Measures and Findings

Table 3 summarizes the sample size, type of evaluation, study type and design, and primary outcomes, and a more detailed description of the main findings for each study is presented in Multimedia Appendix 1 (see Table S1). The sample sizes ranged from 7 to 149,329 participants (median 163, mean 5155). Most studies were formative (20/39, 51%) or process (11/39, 28%) evaluations, with fewer outcome evaluation studies (8/39, 21%). Just over one-third of the studies (14/39, 36%) assessed the health of children or young people as either a primary or secondary outcome; however, only a single (1/14, 7%) study assessed whether using the digital tool improved the health of children or young people [43]. This study examined whether the use of *Mobiletype*, an eHealth tool that allowed general practitioners and young people to monitor symptoms of mood, stress, and daily activities in general practice, was associated with improved mental health outcomes compared with treatment as usual. The authors found that use of the device was associated with a significant improvement in emotional self-awareness but found no changes in symptoms of depression, anxiety, or stress; post hoc analyses showed enhanced mental health care at the initial assessment among general practitioners using the tool compared with those who did not.

XSL•FO
RenderX

Table 3. Outcomes of studies.

Study	Sample (N)	Type of evaluation[a]	Study type	Study design[b]	Health as outcome[c]	Primary outcome	Challenges or risks of using the tools
Alawna et al, 2019 [50]	58	Formative	Quantitative	Descriptive study	No	Reliability (intra- and inter-rater reliability)	• Clinical utility: questionable accuracy of readings in people with certain health conditions (eg, obesity and limb deformity)
Binotti et al, 2019 [56]	40	Formative	Quantitative	Descriptive study	No	Concordance rating[d]	• Clinical safety: partial overestimation of heart rate when <60 beats per minute
Boyce et al, 2019 [48]	799	Process	Mixed methods	Quasi-experimental	No	Efficacy	• Accessibility: hardware and software issues (eg, uploading data) • Uptake: time consuming
Den Boer et al, 2018 [18]	653	Formative	Mixed methods	Descriptive study	Yes	Usability and efficacy	• Accessibility: slow internet connection • UX[e]: buttons lacked visual response to input • Sociocultural: parents or carers said questions about smoking for children aged 6-11 years were inappropriate and insulting
Detsomboonrat and Pisarnturakit 2019 [58]	441	Formative	Quantitative	Descriptive study	No	Acceptability and efficacy	• Accessibility: poor internet connection for some users
Dexheimer et al, 2014 [19]	13,896	Outcome	Quantitative	RCT[f]	No	Efficacy (time from triage to clinical decision)	• Clinical utility: clinicians were already implementing best practice guidelines and conducting education without the eHealth tool
Eikelboom et al, 2005 [5]	66	Formative	Quantitative	Descriptive study	No	Concordance rating	• Data quality: poor image quality • Clinical safety: using eHealth tool alone (without input from a qualified clinician) could result in inaccurate diagnosis and treatment
Estai et al, 2016 [57]	126	Formative	Quantitative	Descriptive study	No	Concordance rating	• Data quality: poor image quality

XSL•FO
RenderX

Study	Sample (N)	Type of evaluation[a]	Study type	Study design[b]	Health as outcome[c]	Primary outcome	Challenges or risks of using the tools
Finocchario-Kessler et al, 2015 [6]	NR[g]	Outcome	Mixed methods	Cross-sectional study	No	Feasibility and efficacy	• Accessibility: slow internet connection in some regions • Language proficiency or literacy: some users unable to use the tool because of low literacy levels • Privacy: concerns about the privacy of data • Clinical safety: high turnover of health care workers requiring continuous retraining of staff or risk of inaccurate use of the tool
Franke et al, 2018 [20]	237	Process	Quantitative	Cross-sectional study	No	Concordance rating	• Clinical utility: data only entered by parent or caregiver and mostly in binary (yes or no) format; information from clinician said to be important but not possible as multi-informant assessment not available
Galvez et al, 2017 [52]	1252	Process	Quantitative	Descriptive study	No	Use and uptake	• Accessibility: only available in countries with internet access and where Google was not blocked
Ginsburg et al, 2015 [16]	7	Formative	Mixed methods	Descriptive study	No	Usability and acceptability	• UX: buttons difficult to navigate, pop-ups distracting, difficulty launching application and recording results, too text heavy or more images needed
Gregory et al, 2017 [39]	76	Formative	Quantitative	Descriptive study	No	Feasibility of uptake	• Uptake: lower than expected uptake by young people
Han et al, 2019 [53]	150	Outcome	Quantitative	Cohort study	Yes	Validity and reliability	• Data quality: mobile phones with low resolution may not clearly show results
Hashemi et al, 2017 [7]	986	Outcome	Quantitative	Descriptive study	Yes	Feasibility	• Clinical safety: efficacy of the tool in screening for psychological symptoms not yet validated
Heida et al, 2018 [62]	170	Outcome	Mixed methods	RCT	Yes	Efficacy	• Uptake: clinicians not adequately prepared for changes in traditional ways of working and reluctant to enter data twice
Hussey and Flynn, 2019 [41]	56	Formative	Mixed methods	Comparative study with historical control group	No	Use and efficacy	

XSL•FO
RenderX

Study	Sample (N)	Type of evaluation[a]	Study type	Study design[b]	Health as outcome[c]	Primary outcome	Challenges or risks of using the tools
							• UX: many features needing improvement (eg, emergency alert button, survey tool, SMS text messaging, and notifications)
Iorfino et al, 2017 [8]	232	Process	Quantitative	Nonrandomized experimental trial	Yes	Efficacy	• Clinical utility or safety: efficacy for individuals with low to moderate suicidality not studied
Jeong et al, 2020 [40]	13	Formative	Mixed methods	Descriptive study	No	Feasibility, acceptability, and usability	• Accessibility: health professionals unable to use the tool because of inadequate training
Jiam et al, 2017 [66]	7	Process	Qualitative	Descriptive study	No	Usability	• Language proficiency or literacy: information beyond children's comprehension capacity and literacy levels
Kassam-Adams et al, 2019 [42]	167	Process	Quantitative	Descriptive study	No	Acceptability and efficacy	• Accessibility: lower-income families could not use the tool because of the cost of mobile data
Kim et al, 2019 [60]	149,329	Process	Mixed methods	Descriptive study	No	Uptake, usability, and efficacy	• Uptake: of the 3 countries where the tool was implemented, uptake was only seen in Korea and not China or Japan • Accessibility: only users with a smartphone could use the eHealth tool • Clinical utility: question as to whether increased rates of influenza signaled a local outbreak or new interest in using the tool
Li et al, 2019 [63]	137	Outcome	Quantitative	Pseudo-RCT	Yes	Utility and efficacy	• NR
March et al, 2018 [17]	18	Formative	Mixed methods	Descriptive study	No	Feasibility and acceptability	• NR
Matin et al, 2020 [59]	18	Formative	Quantitative	Pretest–posttest case series	No	Feasibility and acceptability	• Clinical utility: did not assess parents' accuracy in identifying symptoms aided by the tool; outside of the research study, parents may not receive the same on-call support • Clinical safety: only 1 parent attached the wearable band correctly, leading to many incorrect recordings • UX: device lacked notifications to encourage care seeking when necessary
	3805	Formative	Mixed method	Descriptive study	No		• NR

XSL•FO
RenderX

Study	Sample (N)	Type of evaluation[a]	Study type	Study design[b]	Health as outcome[c]	Primary outcome	Challenges or risks of using the tools
McCulloh et al, 2018 [49]						Use or uptake and usability	
Mohammed et al, 2018 [9]	1446	Formative	Quantitative	Descriptive study	Yes	Feasibility and concordance ratings	• Accessibility: poor internet connection in some areas; low ownership of mobile phones • Data quality: incomplete data entered by some parents
Padidar et al, 2019 [64]	113	Formative	Quantitative	Descriptive study	Yes	Efficacy (concordance rating)	• NR
Rath et al, 2018 [45]	405	Formative	Mixed methods	Descriptive study	Yes	Usability and efficacy	• Clinical utility: the anonymity of users prevented verification of health conditions and initiation of follow-up care
Rath et al, 2019 [65]	1615	Formative	Quantitative	Cohort study	Yes	Efficacy	• NR
Reid et al, 2011 [43]	163	Outcome	Quantitative	RCT	Yes	Change in mental health status	• NR
Singh et al, 2017 [51]	16,490	Process	Quantitative	Descriptive study	Yes	Feasibility	• Language proficiency or literacy: many parents could not read English messages (Hindi translations integrated to address this issue) • Uptake: clinicians and parents were initially resistant to use the new digital system • Data quality: errors in data entry related to free text input • Clinical utility: customization of question sets needed depending on user characteristics
Svedberg et al, 2019 [46]	46	Process	Qualitative	Descriptive study	No	Feasibility and acceptability	• Uptake: low uptake because of required organizational restructuring and competing workplace demands (eg, high workload) • UX: software issues related to printing reports and unwanted termination of sessions
Thabrew et al, 2019 [44]	129	Formative	Mixed methods	Pseudo-RCT	No	Efficacy and acceptability	• Accessibility: some internet connection issues • Language proficiency or literacy: information beyond the comprehension and literacy levels of some low socioeconomic groups
Thabtah, 2018 [54]	1452	Outcome	Quantitative	Descriptive study	Yes	Feasibility and efficacy	• NR

XSL•FO
RenderX

Study	Sample (N)	Type of evaluation[a]	Study type	Study design[b]	Health as outcome[c]	Primary outcome	Challenges or risks of using the tools
Thompson et al, 2016 [47]	937	Process	Quantitative	Descriptive study	No	Use and uptake	• NR
Valdes-Angues et al, 2018 [55]	326	Process	Mixed methods	Descriptive study	No	Feasibility	• Accessibility: poor internet connection; power cuts; inability to recharge device; slow upload speed of data • Data quality: errors in data entry
van Karnebeek et al, 2012 [61]	15	Formative	Qualitative	Descriptive study	Yes	Feasibility and acceptability	• Clinical utility: additional features needed to add value to standard care (eg, entering differential diagnosis and accessing databases with medical information)
Wang et al, 2017 [38]	31	Formative	Qualitative	Descriptive study	No	Usability	• NR

[a]Type of evaluation defined as follows: (1) formative evaluation: assessed feasibility, appropriateness, or acceptability of the digital device before full implementation; (2) process evaluation: assessed whether the digital device had been implemented as intended; (3) outcome evaluation: measured the effectiveness of the digital device by assessing progress in primary outcomes [33].

[b]Study design based on the National Health and Medical Research Council guidelines: randomized controlled trials (RCTs); pseudo-RCTs; comparative studies with concurrent controls, including nonrandomized experimental trials, cohort studies, case–control studies, or interrupted time series with a control group; comparative studies without a control group, including historical control studies, ≥2 single-arm studies or interrupted time series without a parallel control; case series with either posttest or pre- and posttest outcomes; descriptive studies; or other [34].

[c]Whether an aspect of the child's or young person's health was measured as a primary or secondary outcome of the study.

[d]The amount of agreement between the digital tool and clinician ratings.

[e]UX: user experience (user interface or design aspects of the device).

[f]RCT: randomized controlled trial.

[g]NR: not reported.

Challenges and Risks Identified in Studies

Table 3 summarizes the challenges and risks of using the tools identified in each study. Of the studies examined, most (30/39, 77%) identified at least one challenge or risk, which was related to accessibility (11/39, 28%), clinical utility (9/39, 23%) or clinical safety (5/39, 13%) of the tool, uptake by users (6/39, 15%), data quality (6/39, 15%), user interface or design aspects of the device (user experience; 5/39, 13%); language proficiency or literacy barriers (4/39, 10%), sociocultural barriers (1/39, 3%), and privacy concerns (1/39, 3%). More specifically, accessibility problems were related to poor internet connection, inability to recharge devices because of power cuts, slow or inefficient upload of information, lack of access to a device, and low technological literacy of end users. Clinical utility and clinical safety concerns were related to the validity of data among people with different health conditions, lack of appropriate training of staff, input from a health care professional rather than entirely self-report data to ensure safe and accurate interpretation of results, whether the tool added value over and above standard clinical care, and the safety of tools that had not yet been validated to detect clinical symptoms. Uptake of tools was a frequently cited barrier; however, there was often no further investigation or explanation as to why uptake was lower than expected. Data quality concerns were centered on inaccurate or incomplete data entry (because of human or computer error) and poor-quality images. User experience or design barriers referred to the eHealth tool lacking the necessary features to make it functional and usable for end users. Language proficiency or literacy barriers were centered on users lacking the comprehension and literacy levels to understand and take action from the presented information; this was a concern reported when end users were children, non–English speaking, or from lower socioeconomic backgrounds. Sociocultural barriers were related to the appropriateness of questions and the risk of causing offense or harm.

Research Funding

Studies were financially supported by the public sector (ie, government, universities, research institutes, and professional associations) and commercial or NFP organizations (28/39, 72%). Receiving funding from ≥1 sector was the most common (16/39, 41%), followed by funding solely from public sources (6/39, 15%) and NFPs (5/39, 13%). No study was funded solely by the commercial sector; however, commercial funding contributed to nearly one-third of studies with combined funding sources (5/16, 31%). The remainder of the studies did not receive external financial support or did not report it in the paper (11/39, 28%).

XSL•FO
RenderX

Quality Ratings of Selected Papers

The methodological quality of the Downs and Black checklist was rated for 95% (35/39) of studies that included quantitative data: 64% (25/39) of studies had a low chance of bias, 36% (14/39) of studies had a moderate chance of bias, and no studies had a high chance of bias (see Table S2, Multimedia Appendix 1). The National Institute for Health and Care Excellence Quality Appraisal Checklist was completed for 44% (17/39) of studies that included qualitative data: 59% (10/17) of studies received a maximum score of 2 for quality, and 41% (7/17) of studies received a partial score of 1; no studies received a score of 0 (see Table S3, Multimedia Appendix 1).

Discussion

Principal Findings

The aim of this systematic review was to summarize eHealth tools designed to assess and track health outcomes in children and young people to clarify the current scope, nature, and efficacy of this emerging type of technology in health care. Our findings revealed exponential growth in the development and evaluation of these tools over the past 10 years; however, the results showed that the research is still in its infancy, with most studies assessing feasibility, acceptability, usability, or uptake of a device rather than the efficacy of tools in relation to health outcomes. Overall, the current tools showed potential to enhance the assessment and tracking of children or young people in health services around the world. Further research is needed to evaluate the efficacy of tools for improving health outcomes and clinical care delivery, as well as to identify and address the risks and challenges of implementing these tools as part of standard clinical care.

There are numerous potential advantages of using eHealth solutions for children and young people, including the ability to conduct multidomain and multi-informant assessments, undertake continuous monitoring, and assist with timely connection to personalized clinical care [1,2,13]. Encouragingly, over half of the tools facilitated a connection between the child or young person and a health care provider; however, less than half tracked children's health data over time or collected information from multiple informants (ie, child or young person, parent or caregiver, and health care professional). These findings demonstrate an untapped potential of eHealth solutions in facilitating multi-informant assessments and longitudinally tracking health over time among children or young people, which is key to achieving comprehensive, multidisciplinary care [2]. In addition, data were most commonly collected in the form of surveys or questionnaires, illustrating a lag in uptake and integration of newer technologies (eg, biosensors or wearables to collect physiological data). Such technology has the potential to enhance symptom detection and clinical decision-making [13,14,19,48,49] and may be an important area for future research to explore.

Another potential advantage of eHealth is its ability to overcome geographical, financial, and social barriers that hinder the provision of health services in specific populations and locations [10,67]. A small number of studies evaluated devices that were implemented across multiple countries [45,52-55], highlighting the ability of digital technologies to provide health care with greater reach. However, fewer tools were implemented in rural areas compared with urban areas, and there was less implementation in community outreach settings compared with hospitals and health clinics. Although using eHealth tools in health clinics and hospitals is a step forward from traditional paper-based methods in terms of data management and integrated care, there is greater potential for the tools to engage hard-to-reach populations in regional and community settings [9,55]. The higher percentage of devices used in health clinics and hospitals may be as community settings do not always incorporate systematic health tracking into their procedures or reflect a lack of availability of skilled health professionals to collect and enter health information in community settings. A number of studies overcame this issue by using trained health workers (ie, staff who received specific training in the health condition and digital device but were not specialists in the field); these trained health workers were able to collect information in the community, with studies finding that this did not compromise the validity or reliability of data or clinical care [6,7,48,50,56,57]. Another solution was to collect information solely from the consumer (ie, child or young person or their parent or guardian), which is particularly common in studies examining mental health [8,43,44]. Together, these results demonstrate room for broader implementation in rural or remote regions and community settings around the world. Nevertheless, an important issue to recognize is that rural and vulnerable populations are currently experiencing the largest digital divide [68-70]. Inequalities in access exist because of variations in location, age, education, and income level. For instance, the cost of internet access is higher in rural or remote regions than in urban areas, which is compounded by the fact that some rural residents have less disposable income than their metropolitan counterparts. Thus, to truly overcome geographical, financial, and social barriers and reach these populations, researchers must consider the broader socioeconomic context from which these access issues stem.

The eHealth tools studied focused on various domains of health, including infectious diseases, child development, and neurological conditions; however, the most widely assessed domains were general and mental health. The focus on mental health tools may reflect a growing need and demand for mental health care among children and young people, increasing evidence supporting eHealth in the field of mental health, and increased funding for mental health tools [71,72]. Only one of the eHealth tools assessed multiple health domains [38], despite this often being important for gaining a holistic picture of a child or young person's health concerns. The development of eHealth tools that assess health multidimensionally is likely to be important in future eHealth tools, perhaps assessing not only current symptoms but also broader social or environmental factors related to the etiology and trajectory of illness and barriers to or facilitators of accessing care [10,13].

Various challenges and risks were identified in relation to the implementation and use of eHealth tools. These barriers were related to the accessibility and functionality of devices, including poor internet connection [6,9,15,18,44,48,52,55,58] and user interface or design aspects of the tool [16,18,41,46,59]. Clinical

XSL•FO
RenderX

utility was another barrier identified in papers, mainly relating to the tool lacking features that provided added value to standard care [8,19,45,51,59-61]. User uptake was a challenge, with multiple studies reporting lower than expected uptake and reluctance to use the digital tool; this was an issue reported among diverse user groups, including clinicians, parents or carers, and children or young people [39,46,48,51,59,60,62]. These findings are in line with the Eysenbach [73] law of attrition, which is based on the observation that high participant dropout rates are common in eHealth research focusing on novel digital health tools; although researchers may dismiss or underreport this information, the observation meaningfully reflects the real-world uptake of digital tools currently. Some of the reasons for low uptake included implementation barriers (eg, competing time, modified professional roles, and organizational restructuring) [46,48,62], privacy concerns [6], socioeconomic factors (eg, cost of data) [42], and language proficiency or literacy issues [6,44,51]. Incorporating participatory design (co-design) and user testing methodologies into future protocols may help to understand and address these barriers [10,22]. Data quality was another barrier that was reported, which was related to human error in data entry [51,55] or incomplete data input [9]. A study overcame the issue of human error by minimizing free-text input and using predefined options [51]. Although this is not feasible for all tools, such as when obtaining qualitative health information, it provides a solution for quantitative health data. Sociocultural issues were mentioned in just 1 study; Den Boer [18] reported that parents or carers in some communities found questions about smoking in children aged 6 to 11 years *insulting* and inappropriate. The study researchers justified the inclusion of the questions by saying that they were important and relevant for certain communities or user groups. This raises the issue of whether universal questionnaires can be used in eHealth tools or whether customized question sets need to be developed for the target group. The findings of Singh [51] supported a configurable or individualized approach, with the researchers concluding that individualization was critical to the clinical utility and safety of eHealth tools. Despite studies identifying challenges and risks, there was minimal discussion on how to address the identified issues; moreover, just under a quarter of studies did not report any potential risks or challenges of implementing the device [17,38,43,47,49,54,63-65].

Our review of funding sources, which showed that financial support came from a mixture of public, commercial, and NFP bodies, is unsurprising, as the development and implementation of digital tools often involve the collaboration of professionals from multiple disciplines who belong to different bodies [25,74]. Interestingly, funding from commercial bodies was uncommon. This may reflect the fact that the tools reviewed were in the preliminary stages of research (development and implementation) rather than at a more advanced stage of commercialization, the latter of which we would expect to attract more investment from commercial organizations [74].

Recommendations for Future Research

The findings of this review demonstrate a clear need for further research into the efficacy and validity of eHealth tools that assess and track health outcomes in children and young people. Future evaluation studies should focus on changes in the health outcomes of users, as well as the clinical care pathways. Further investigation of the risks and challenges of implementing devices is also important, particularly relating to sociocultural factors, language proficiency or literacy, and privacy concerns, as these were seldom mentioned but are likely to affect the clinical utility, safety, and uptake of tools [10,23]. Overall, these findings are consistent with results from a prior systematic review of eHealth solutions in adults, which found a gap between the postulated and empirically demonstrated benefits of eHealth technologies, a lack of robust research trials into validity and efficacy, and inadequate investigation of risks or challenges of using these technologies in health care [75]. This review has uncovered several features of eHealth tools that may facilitate comprehensive assessments and integrated care in future technologies:

1. Capacity for multi-informant assessment, including input from a health professional and the child or young person or their parent or caregiver
2. Multidomain assessments, allowing for a holistic picture of the child or young person's health to be captured rather than assessing health in one domain
3. Tracking over time (ie, capacity and use of tools for assessment at multiple time points)
4. Configurability of question sets or content depending on characteristics (eg, demographic, sociocultural, and health concerns) of the target group
5. Connection to clinical care that is tailored to the child or young person's current needs
6. Trialing integration of newer technologies (eg, biosensors or wearables to collect physiological data) for relevant health domains

Limitations

Although this review provides important insights into a novel field of eHealth, the conclusions that can be drawn about the efficacy and validity of eHealth solutions are limited as most studies were formative and process evaluations that assessed feasibility, acceptability, usability, or uptake of a device. Outcome evaluation studies were rare, with just 3 randomized controlled trials conducted to date. These early phase research studies are necessary precursors to more rigorous validity and efficacy studies; however, they need to be followed by more thorough evaluation studies to determine whether the tools are effective in improving health outcomes and clinical care. We limited our search to studies published in English, which may have biased our results. Furthermore, although the strength of this review is that it presents the state of eHealth tools for supporting health in children or young people, it inevitably fails to consider the immense variation that lies within each health domain. Our search strategy was not without limitations. We did not include all relevant terms (eg, internet and technology) as the inclusion of these broader terms returned >15,000 articles, which was not considered realistic for screening. Nevertheless, we believe that the search strategy balanced scientific rigor and feasibility and was sufficiently rigorous to pick up relevant articles. Finally, this paper was not preregistered with PROSPERO; however, the search strategy remained the same over time.

XSL•FO
RenderX

Conclusions

eHealth tools that assess and track health outcomes in children or young people and connect individuals with personalized care options have enormous potential in health services around the world. Many of the existing tools are in the early stages of pilot and feasibility testing; however, the literature is promising in the potential to use these tools in future clinical care. Further research is needed to evaluate the validity and efficacy of these eHealth tools and investigate the potential risks and challenges of implementation as part of standard clinical care. With future research and development efforts in place, these tools have the potential to facilitate collaborative decision-making, improved communication, transmission of remote health data, and real-time assessment and tracking and take a positive step forward in digitalizing health practices.

Conflicts of Interest

TD is now Director, Research & Insights at the Australian Digital Health Agency.

Multimedia Appendix 1
Digital review.
[DOCX File , 72 KB - jmir_v24i5e26015_app1.docx]

References

1. UNICEF Annual Report 2017. UNICEF. 2018. URL: https://www.unicef.org/reports/unicef-annual-report-2017 [accessed 2021-12-19]
2. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Early Childhood Development Interventions Review Group, for the Lancet Early Childhood Development Series Steering Committee. Nurturing care: promoting early childhood development. Lancet 2017 Jan 07;389(10064):91-102. [doi: 10.1016/S0140-6736(16)31390-3] [Medline: 27717615]
3. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
4. 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 Apr;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
5. Eikelboom RH, Mbao MN, Coates HL, Atlas MD, Gallop MA. Validation of tele-otology to diagnose ear disease in children. Int J Pediatr Otorhinolaryngol 2005 Jun;69(6):739-744. [doi: 10.1016/j.ijporl.2004.12.008] [Medline: 15885325]
6. Finocchario-Kessler S, Odera I, Okoth V, Bawcom C, Gautney B, Khamadi S, et al. Lessons learned from implementing the HIV infant tracking system (HITSystem): a web-based intervention to improve early infant diagnosis in Kenya. Healthc (Amst) 2015 Dec;3(4):190-195 [FREE Full text] [doi: 10.1016/j.hjdsi.2015.07.004] [Medline: 26699342]
7. Hashemi B, Ali S, Awaad R, Soudi L, Housel L, Sosebee SJ. Facilitating mental health screening of war-torn populations using mobile applications. Soc Psychiatry Psychiatr Epidemiol 2017 Jan;52(1):27-33. [doi: 10.1007/s00127-016-1303-7] [Medline: 27815623]
8. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: a cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247 [FREE Full text] [doi: 10.2196/jmir.7897] [Medline: 28701290]
9. Mohammed A, Franke K, Okyere P, Brinkel J, Marinovic A, Kreuels B, et al. Feasibility of Electronic Health Information and Surveillance System (eHISS) for disease symptom monitoring: a case of rural Ghana. PLoS One 2018 May 24;13(5):e0197756 [FREE Full text] [doi: 10.1371/journal.pone.0197756] [Medline: 29795626]
10. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:3-39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
11. Davenport TA, LaMonica HM, Whittle L, English A, Iorfino F, Cross S, et al. Validation of the InnoWell platform: protocol for a clinical trial. JMIR Res Protoc 2019 May 31;8(5):e13955 [FREE Full text] [doi: 10.2196/13955] [Medline: 31152524]
12. Rowe SC, Davenport TA, Easton MA, Jackson TA, Melsness J, Ottavio A, et al. Co-designing the InnoWell Platform to deliver the right mental health care first time to regional youth. Aust J Rural Health 2020 Apr;28(2):190-194. [doi: 10.1111/ajr.12617] [Medline: 32281183]
13. Hickie IB, Scott EM, Cross SP, Iorfino F, Davenport TA, Guastella AJ, et al. Right care, first time: a highly personalised and measurement-based care model to manage youth mental health. Med J Aust 2019 Nov;211 Suppl 9:3-46. [doi: 10.5694/mja2.50383] [Medline: 31679171]
14. Iorfino F, Cross SP, Davenport T, Carpenter JS, Scott E, Shiran S, et al. A digital platform designed for youth mental health services to deliver personalized and measurement-based care. Front Psychiatry 2019 Aug 23;10:595 [FREE Full text] [doi: 10.3389/fpsyt.2019.00595] [Medline: 31507465]

XSL•FO
RenderX

15. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
16. Ginsburg AS, Delarosa J, Brunette W, Levari S, Sundt M, Larson C, et al. mPneumonia: development of an innovative mHealth application for diagnosing and treating childhood pneumonia and other childhood illnesses in low-resource settings. PLoS One 2015 Oct 16;10(10):e0139625 [FREE Full text] [doi: 10.1371/journal.pone.0139625] [Medline: 26474321]
17. March S, Day J, Zieschank K, Ireland M. The interactive child distress screener: development and preliminary feasibility testing. JMIR Mhealth Uhealth 2018 Apr 19;6(4):e90 [FREE Full text] [doi: 10.2196/mhealth.9456] [Medline: 29674310]
18. den Boer JC, van Dijk W, Horn V, Hescot P, Bruers JJ. Collecting standardised oral health data via mobile application: a proof of concept study in the Netherlands. PLoS One 2018 Feb 7;13(2):e0191385 [FREE Full text] [doi: 10.1371/journal.pone.0191385] [Medline: 29415025]
19. Dexheimer JW, Abramo TJ, Arnold DH, Johnson K, Shyr Y, Ye F, et al. Implementation and evaluation of an integrated computerized asthma management system in a pediatric emergency department: a randomized clinical trial. Int J Med Inform 2014 Nov;83(11):805-813 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.07.008] [Medline: 25174321]
20. Franke KH, Krumkamp R, Mohammed A, Sarpong N, Owusu-Dabo E, Brinkel J, et al. A mobile phone based tool to identify symptoms of common childhood diseases in Ghana: development and evaluation of the integrated clinical algorithm in a cross-sectional study. BMC Med Inform Decis Mak 2018 Mar 27;18(1):23 [FREE Full text] [doi: 10.1186/s12911-018-0600-3] [Medline: 29580278]
21. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
22. LaMonica HM, Davenport TA, Braunstein K, Ottavio A, Piper S, Martin C, et al. Technology-enabled person-centered mental health services reform: strategy for implementation science. JMIR Ment Health 2019 Sep 19;6(9):e14719 [FREE Full text] [doi: 10.2196/14719] [Medline: 31538938]
23. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-enabled mental health service reform for open arms - veterans and families counselling: participatory design study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
24. Ospina-Pinillos L, Davenport T, Diaz AM, Navarro-Mancilla A, Scott EM, Hickie IB. 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 Aug 02;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
25. Eng TR. eHealth research and evaluation: challenges and opportunities. J Health Commun 2002;7(4):267-272. [doi: 10.1080/10810730290001747] [Medline: 12356287]
26. National Safety and Quality Digital Mental Health (NSQDMH) standards. Australian Commission on Safety and Quality in Healthcare. 2020. URL: https://www.safetyandquality.gov.au/sites/default/files/2020-04/UPDATED%20national_safety_and_quality_digital_mental_health_standards_-_consultation_draft_-_feb_2020_0.pdf [accessed 2021-12-19]
27. Draft global strategy on digital health 2020–2025. World Health Organization. 2020. URL: https://www.who.int/docs/default-source/documents/gs4dhdaa2a9f352b0445bafbc79ca799dce4d.pdf?sfvrsn=f112ede5_42 [accessed 2021-12-19]
28. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
29. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011;17(6):484-494. [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
30. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
31. Adolescent health and development. World Health Organization. 2019. URL: https://tinyurl.com/232vmuy4 [accessed 2021-12-19]
32. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
33. Types of evaluation. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/std/Program/pupestd/Types%20of%20Evaluation.pdf [accessed 2021-12-19]
34. NHMRC additional levels of evidence and grades for recommendations for developers of guidelines. Australian Government National Health and Medical Research Council. 2008. URL: https://www.mja.com.au/sites/default/files/NHMRC.levels.of.evidence.2008-09.pdf [accessed 2021-12-19]
35. Dijkstra L, Florczyk A, Freire S, Kemper T, Melchiorri M, Pesaresi M, et al. Applying the Degree of Urbanisation to the globe: a new harmonised definition reveals a different picture of global urbanisation. J Urban Econ 2021 Sep;125:103312 [FREE Full text] [doi: 10.1016/j.jue.2020.103312]
36. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998 Jun;52(6):377-384 [FREE Full text] [doi: 10.1136/jech.52.6.377] [Medline: 9764259]

XSL•FO
RenderX

37. Methods for the development of NICE public health guidance. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/process/pmg4/chapter/appendix-h-quality-appraisal-checklist-qualitative-studies [accessed 2021-12-19]
38. Wang J, Yao NA, Liu Y, Geng Z, Wang Y, Shen N, et al. Development of a smartphone application to monitor pediatric patient-reported outcomes. Stud Health Technol Inform 2017;245:253-257. [Medline: 29295093]
39. Gregory JM, Sukhera J, Taylor-Gates M. Integrating smartphone technology at the time of discharge from a child and adolescent inpatient psychiatry unit. J Can Acad Child Adolesc Psychiatry 2017;26(1):45-50 [FREE Full text] [Medline: 28331503]
40. Jeong Y, Chang H, Kim J. Development and feasibility of a safety plan mobile application for adolescent suicide attempt survivors. Comput Inform Nurs 2020 Aug;38(8):382-392. [doi: 10.1097/CIN.0000000000000592] [Medline: 32079816]
41. Hussey D, Flynn KC. The utility and impact of the addiction comprehensive health enhancement support system (ACHESS) on substance abuse treatment adherence among youth in an intensive outpatient program. Psychiatry Res 2019 Nov;281:112580. [doi: 10.1016/j.psychres.2019.112580] [Medline: 31627070]
42. Kassam-Adams N, Kohser KL, McLaughlin J, Winston F, Marsac ML. Evaluating the acceptability and validity of assessing pain and posttraumatic stress symptoms in an adaptable eHealth system for school-age children. Clin Pract Pediatr Psychol 2019 Mar;7(1):9-19 [FREE Full text] [doi: 10.1037/cpp0000261] [Medline: 31275781]
43. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011 Nov 29;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
44. Thabrew H, D'Silva S, Darragh M, Goldfinch M, Meads J, Goodyear-Smith F. Comparison of YouthCHAT, an electronic composite psychosocial screener, with a clinician interview assessment for young people: randomized controlled trial. J Med Internet Res 2019 Dec 03;21(12):e13911 [FREE Full text] [doi: 10.2196/13911] [Medline: 31793890]
45. Rath B, Swenshon S, Haase K, Szeles T, Jung C, Jacobi F, et al. Using a mobile application to detect health needs among children and adolescents who are newly arrived migrants in Europe. J Public Health (Oxf) 2019 Dec 20;41(4):840-849. [doi: 10.1093/pubmed/fdy191] [Medline: 30423143]
46. Svedberg P, Arvidsson S, Larsson I, Carlsson I, Nygren JM. Barriers and enablers affecting successful implementation of the electronic health service Sisom: multicenter study of child participation in pediatric care. J Med Internet Res 2019 Nov 15;21(11):e14271 [FREE Full text] [doi: 10.2196/14271] [Medline: 31730040]
47. Thompson LA, Martinko T, Budd P, Mercado R, Schentrup AM. Meaningful use of a confidential adolescent patient portal. J Adolesc Health 2016 Feb;58(2):134-140. [doi: 10.1016/j.jadohealth.2015.10.015] [Medline: 26802988]
48. Boyce S, Nyangara F, Kamunyori J. A mixed-methods quasi-experimental evaluation of a mobile health application and quality of care in the integrated community case management program in Malawi. J Glob Health 2019 Jun;9(1):010811 [FREE Full text] [doi: 10.7189/jogh.09.010811] [Medline: 31263554]
49. McCulloh RJ, Fouquet SD, Herigon J, Biondi EA, Kennedy B, Kerns E, et al. Development and implementation of a mobile device-based pediatric electronic decision support tool as part of a national practice standardization project. J Am Med Inform Assoc 2018 Sep 01;25(9):1175-1182 [FREE Full text] [doi: 10.1093/jamia/ocy069] [Medline: 29889255]
50. Alawna M, Unver B, Yuksel E. The reliability of a smartphone goniometer application compared with a traditional goniometer for measuring ankle joint range of motion. J Am Podiatr Med Assoc 2019 Jan;109(1):22-29. [doi: 10.7547/16-128] [Medline: 30964318]
51. Singh H, Mallaiah R, Yadav G, Verma N, Sawhney A, Brahmachari SK. iCHRCloud: web and mobile based child health imprints for smart healthcare. J Med Syst 2017 Nov 29;42(1):14. [doi: 10.1007/s10916-017-0866-5] [Medline: 29188446]
52. Gálvez JA, Lockman JL, Schleelein LE, Simpao AF, Ahumada LM, Wolf BA, et al. Interactive pediatric emergency checklists to the palm of your hand - How the Pedi Crisis App traveled around the world. Paediatr Anaesth 2017 Aug;27(8):835-840. [doi: 10.1111/pan.13173] [Medline: 28593682]
53. Han X, Scheetz J, Keel S, Liao C, Liu C, Jiang Y, et al. Development and validation of a smartphone-based visual acuity test (vision at home). Transl Vis Sci Technol 2019 Jul;8(4):27 [FREE Full text] [doi: 10.1167/tvst.8.4.27] [Medline: 31440424]
54. Thabtah F. An accessible and efficient autism screening method for behavioural data and predictive analyses. Health Informatics J 2019 Dec;25(4):1739-1755 [FREE Full text] [doi: 10.1177/1460458218796636] [Medline: 30230414]
55. Angues RV, Suits A, Palmer VS, Okot C, Okot RA, Atonywalo C, et al. A real-time medical cartography of epidemic disease (Nodding syndrome) using village-based lay mHealth reporters. PLoS Negl Trop Dis 2018 Jun;12(6):e0006588 [FREE Full text] [doi: 10.1371/journal.pntd.0006588] [Medline: 29906291]
56. Binotti M, Cavallin F, Ingrassia PL, Pejovic NJ, Monzani A, Genoni G, et al. Heart rate assessment using NeoTapAdvancedSupport: a simulation study. Arch Dis Child Fetal Neonatal Ed 2019 Jul;104(4):440-442. [doi: 10.1136/archdischild-2018-315408] [Medline: 30262463]
57. Estai M, Winters J, Kanagasingam Y, Shiikha J, Checker H, Kruger E, et al. Validity and reliability of remote dental screening by different oral health professionals using a store-and-forward telehealth model. Br Dent J 2016 Oct 07;221(7):411-414. [doi: 10.1038/sj.bdj.2016.733] [Medline: 27713449]

XSL•FO
RenderX

58. Detsomboonrat P, Pisarnturakit PP. Development and evaluation: the satisfaction of using an oral health survey mobile application. Telemed J E Health 2019 Jan;25(1):55-59. [doi: 10.1089/tmj.2017.0288] [Medline: 29870315]
59. Matin SB, Wallingford A, Xu S, Ng N, Ho A, Vanosdoll M, et al. Feasibility of a mobile health tool for mothers to identify neonatal illness in rural Uganda: acceptability study. JMIR Mhealth Uhealth 2020 Feb 18;8(2):e16426 [FREE Full text] [doi: 10.2196/16426] [Medline: 32130174]
60. Kim M, Yune S, Chang S, Jung Y, Sa SO, Han HW. The fever coach mobile app for participatory influenza surveillance in children: usability study. JMIR Mhealth Uhealth 2019 Oct 17;7(10):e14276 [FREE Full text] [doi: 10.2196/14276] [Medline: 31625946]
61. van Karnebeek CD, Houben RF, Lafek M, Giannasi W, Stockler S. The treatable intellectual disability APP www. treatable-id. org: a digital tool to enhance diagnosis and care for rare diseases. Orphanet J Rare Dis 2012 Jul 23;7:47 [FREE Full text] [doi: 10.1186/1750-1172-7-47] [Medline: 22824307]
62. Heida A, Dijkstra A, Kobold AM, Rossen J, 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]
63. Li C, Huang S, Su X, Zhang T, Jiang K. Monitoring of home recovery using the 317-nursing mobile application following day-case surgery in children: perspectives from both nurses and patients. Medicine (Baltimore) 2019 Aug;98(31):e16639 [FREE Full text] [doi: 10.1097/MD.0000000000016639] [Medline: 31374033]
64. Padidar P, Shaker M, Amoozgar H, Khorraminejad-Shirazi M, Hemmati F, Najib KS, et al. Detection of neonatal jaundice by using an android OS-based smartphone application. Iran J Pediatr 2019 Feb 02 (forthcoming). [doi: 10.5812/ijp.84397]
65. Rath B, Maltezou HC, Papaevangelou V, Papagrigoriou-Theodoridou MA, Alchikh M, Myles P, PEDSIDEA Network. Partnering for enhanced digital surveillance of influenza-like disease and the effect of antivirals and vaccines (PEDSIDEA). Influenza Respir Viruses 2019 Jul;13(4):309-318 [FREE Full text] [doi: 10.1111/irv.12645] [Medline: 31169347]
66. Jiam NT, Hoon AH, Hostetter CF, Khare MM. IIAM (important information about me): a patient portability profile app for adults, children and families with neurodevelopmental disabilities. Disabil Rehabil Assist Technol 2017 Aug;12(6):599-604. [doi: 10.1080/17483107.2016.1198435] [Medline: 27377240]
67. The rural and remote implications of a national e-health strategy. National Health Rural Alliance. 2009. URL: https://www.ruralhealth.org.au/sites/default/files/position-papers/position-paper-09-07-23.pdf [accessed 2021-12-19]
68. Curtin J. A digital divide in rural and regional Australia? Current Issues Briefs - Parliament of Australia. 2001. URL: https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/Publications_Archive/CIB/cib0102/02CIB01#top [accessed 2021-12-19]
69. Henry L. Bridging the urban-rural digital divide and mobilizing technology for poverty eradication: challenges and gaps. GSM Association. URL: https://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2019/03/Henry-Bridging-the-Digital-Divide-2019.pdf [accessed 2021-12-19]
70. Walker J. Adding to the list of rural inequalities-Digital technology exclusion. Aust J Rural Health 2019 Oct 30;27(5):440-441. [doi: 10.1111/ajr.12585] [Medline: 31663211]
71. Christensen H, Hickie IB. Using e-health applications to deliver new mental health services. Med J Aust 2010 Jun 07;192(S11):53-56. [doi: 10.5694/j.1326-5377.2010.tb03695.x] [Medline: 20528711]
72. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 07;192(S11):4-11. [doi: 10.5694/j.1326-5377.2010.tb03685.x] [Medline: 20528707]
73. 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] [Medline: 15829473]
74. Wu YP, Steele RG, Connelly MA, Palermo TM, Ritterband LM. Commentary: pediatric eHealth interventions: common challenges during development, implementation, and dissemination. J Pediatr Psychol 2014 Jul;39(6):612-623 [FREE Full text] [doi: 10.1093/jpepsy/jsu022] [Medline: 24816766]
75. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]

Abbreviations

NFP: not-for-profit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Edited by R Kukafka, G Eysenbach; submitted 26.11.20; peer-reviewed by L Ospina-Pinillos, A Čuš; comments to author 14.01.21; revised version received 15.04.21; accepted 14.10.21; published 12.05.22.

Please cite as:
Stewart E, Milton A, Yee HF, Song MJ, Roberts A, Davenport T, Hickie I
eHealth Tools That Assess and Track Health and Well-being in Children and Young People: Systematic Review
J Med Internet Res 2022;24(5):e26015
URL: https://www.jmir.org/2022/5/e26015
doi:10.2196/26015
PMID:35550285

©Elizabeth Stewart, Alyssa Milton, Hannah Frances Yee, Michael Jae Song, Anna Roberts, Tracey Davenport, Ian Hickie. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2022. 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.

XSL•FO
RenderX

Original Paper

Use of Robots in Critical Care: Systematic Review

Rachel Teng[1*]; Yichen Ding[1*]; Kay Choong See[2], MBBS, MPH

[1]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Division of Respiratory & Critical Care Medicine, Department of Medicine, National University Hospital, Singapore, Singapore

[*]these authors contributed equally

Corresponding Author:
Yichen Ding
Yong Loo Lin School of Medicine
National University of Singapore
10 Medical Dr
Singapore, 117597
Singapore
Phone: 65 6772 3737
Email: yichending@u.nus.edu

Abstract

Background: The recent focus on the critical setting, especially with the COVID-19 pandemic, has highlighted the need for minimizing contact-based care and increasing robotic use. Robotics is a rising field in the context of health care, and we sought to evaluate the use of robots in critical care settings.

Objective: Although robotic presence is prevalent in the surgical setting, its role in critical care has not been well established. We aimed to examine the uses and limitations of robots for patients who are critically ill.

Methods: This systematic review was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. MEDLINE, Embase, IEEE Xplore, and ACM Library were searched from their inception to December 23, 2021. Included studies involved patients requiring critical care, both in intensive care units or high-dependency units, or settings that required critical care procedures (eg, intubation and cardiopulmonary resuscitation). Randomized trials and observational studies were included.

Results: A total of 33 studies were included. The greatest application of robots in the intensive care unit was in the field of telepresence, whereby robots proved advantageous in providing a reduced response time, earlier intervention, and lower mortality rates. Challenges of telepresence included regulatory and financial barriers. In therapy and stroke rehabilitation, robots achieved superior clinical outcomes safely. Robotic use in patient evaluation and assessment was mainly through ultrasound evaluation, obtaining satisfactory to superior results with the added benefits of remote assessment, time savings, and increased efficiency. Robots in drug dispensing and delivery increased efficiency and generated cost savings. All the robots had technological limitations and hidden costs.

Conclusions: Overall, our results show that robotic use in critical care settings is a beneficial, effective, and well-received intervention that delivers significant benefits to patients, staff, and hospitals. Looking ahead, it is necessary to form strong ethical and legislative frameworks and overcome various regulatory and financial barriers.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42021234162; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=234162

(J Med Internet Res 2022;24(5):e33380) doi:10.2196/33380

KEYWORDS

COVID-19; intensive care; high dependency; telepresence; intubation

Introduction

Robotics is a rising field in the context of health care [1]. Although there has been a surge in the popularity of automated and semiautomated processes in robotic surgery, little research has been conducted on robotic use outside surgical settings. The recent focus on critical care settings, especially in light of the COVID-19 pandemic, with more patients requiring intensive care, monitoring, and treatment, has accentuated the importance of minimizing contact-based care while ensuring efficiency [2].

XSL•FO
RenderX

With regard to the perception and acceptance of robots by health care workers, the COVID-19 pandemic has certainly emphasized the need for more widespread robotic use.

However, there may be underlying concerns with regard to robot safety and job replacements. We hypothesize that, given the current robotic technology, the benefits of robots may be limited to replacing mundane tasks and that use is limited by logistic, ethical, and financial barriers. Therefore, we aimed to examine the benefits and limitations of robots and uncover any significant applications of robotic technology in the critical care setting.

To better evaluate the use of robots against conventional methods of care in critical care settings, we reviewed randomized controlled trials (RCTs) and observational studies. We hope to provide information that allows clinicians and policy makers to assess various areas affected by robotic use and find an appropriate role for robots within the intensive care setting. In addition, we hope that our findings can stimulate further development of robotic technology, including its combination with artificial intelligence (AI).

Methods

Search Strategy and Selection Criteria

The study has been registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42021234162) and was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [3]. A total of 2 authors (RT and YD) independently and systematically searched PubMed, Embase, IEEE Xplore, and ACM Library for all relevant studies published from inception to December 23, 2021, using the *patient or population, intervention, comparison, and outcomes* search strategy [4]. Multimedia Appendix 1 provides the detailed search strategy (Tables S1-S3 in Multimedia Appendix 1). In addition, other studies were identified by scanning the reference lists of articles. No limits were applied for language. Disagreements were resolved with the senior author (KCS).

Robots are defined as any machine capable of performing a series of actions, either autonomously or with external guidance. Critical care is defined as the care of patients with severe illnesses requiring intensive care, monitoring, and treatment. Studies were included if they were RCTs and observational studies reporting robotic use on human participants in critical care settings (intensive care unit [ICU], burns unit, high-dependency unit, critical care, and neonatal ICU [NICU]) or during procedures required in critical care settings (intubation, ventilation, tracheostomy, cannulation, resuscitation, and dialysis). Articles were excluded if they had an irrelevant topic, wrong patient type (nonhuman participants), or wrong setting (surgical setting). Gray literature (preprint and conference abstracts) was excluded because of incomplete descriptions of the relevant areas.

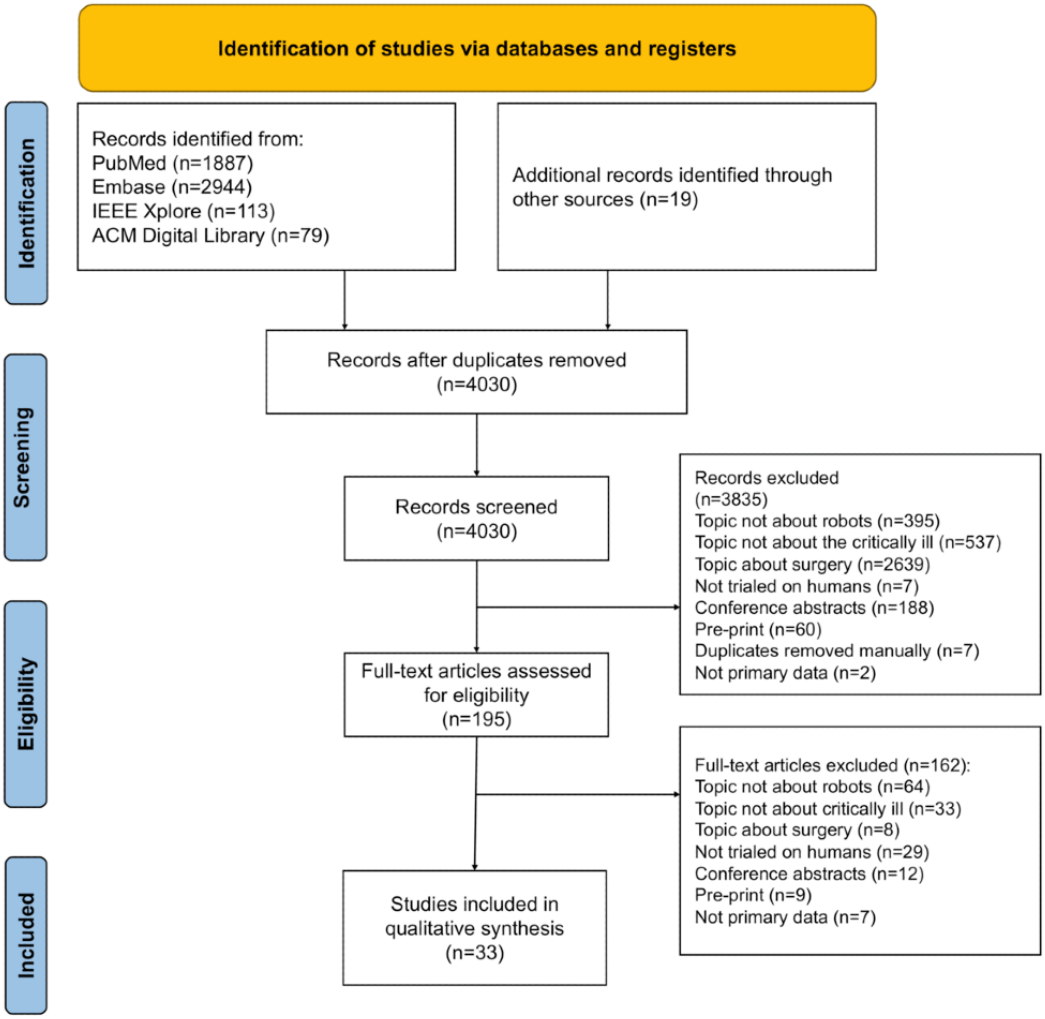
Results

Study Selection

PubMed, Embase, IEEE Xplore, ACM Digital Library, and reference list searches yielded a total of 5042 citations, of which 33 (0.65%) studies were identified for inclusion in the review (Figure 1).

XSL•FO
RenderX

Figure 1. Study flow diagram.



Data Extraction, Quality Assessment, and Data Synthesis

The extracted data included the benefits and limitations of robots. Included studies were independently assessed by 2 authors (RT and YD) for risk of bias using the Standard Quality Assessment Criteria (Tables 1 and 2) [5]. Each study was evaluated based on 14 criteria and scored according to the degree to which the criteria were met (yes, partial, or no). Items not applicable were marked as *N/A* and were excluded from the calculation of the summary score. Disagreements were resolved with the senior author (KCS). The Standard Quality Assessment Criteria suggests a cutoff point of 55% to 75% as an inclusion threshold. Of the 33 included studies, 27 (82%) attained a score of at least 65%. However, we did not exclude studies based on quality scores as this would arbitrarily limit data comprehensiveness.

With regard to data synthesis, given that study designs, participants, interventions, and reported outcomes were expected to vary across papers, we focused on the qualitative synthesis and did not conduct a meta-analysis. We have described the studies in terms of their results, applicability, and limitations.

XSL•FO
RenderX

Table 1. Risk of bias assessment of included studies using Standard Quality Assessment Criteria (study design and interventions).

Study	Study design				Interventions		
	Objective described	Evident and appropriate study design	Participant selection described and appropriate	Participant characteristics described	Random allocation described	Blinding of investigators reported	Blinding of participants reported
Adcock et al [6]	Yes	Yes	Yes	Yes	Yes	N/A[a]	N/A
Alnobani et al [7]	Yes	Yes	Yes	Partial	N/A	N/A	N/A
Amodeo et al [8]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Becevic et al [9]	Partial	Partial	No	Partial	N/A	N/A	N/A
Bettinelli et al [10]	Yes	Yes	Yes	Yes	Partial	N/A	N/A
Burke et al [11]	Partial	Partial	Partial	Partial	N/A	N/A	N/A
Duan et al [12]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Frazzitta et al [13]	Yes	Yes	Yes	Yes	Yes	N/A	N/A
Garingo et al [14]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Garingo et al [15]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Goldberg et al [16]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Holsti et al [17]	Yes	Yes	Yes	Yes	Yes	Yes	N/A
Holt et al [18]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Ito et al [19]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Lazzara et al [20]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Marini et al [21]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Marttos et al [22]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
McNelis et al [23]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Murray et al [24]	No	No	No	No	N/A	N/A	N/A
Prokazova et al [25]	Yes	Yes	Yes	Yes	No	N/A	N/A
Reynolds et al [26]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Rincon et al [27]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Rocca et al [28]	Yes	Yes	Yes	Yes	Yes	N/A	N/A
Rogove et al [29]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Ruiz-Del-Solar et al [30]	Yes	Yes	Yes	Partial	N/A	N/A	N/A
Shimizu et al [31]	Partial	Partial	Partial	Partial	N/A	N/A	N/A
Sucher et al [32]	Yes	Yes	Yes	Partial	N/A	N/A	N/A
Summerfield et al [33]	Yes	Yes	Partial	Yes	N/A	N/A	N/A
Vespa et al [34]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Wang et al [35]	Yes	Yes	Partial	Partial	N/A	N/A	N/A
Williams et al [36]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ye et al [37]	Yes	Yes	Yes	Yes	N/A	N/A	N/A
Zeiler et al [38]	Yes	Yes	Yes	Yes	N/A	N/A	N/A

[a]N/A: not applicable.

XSL•FO
RenderX

Table 2. Risk of bias assessment of included studies using Standard Quality Assessment Criteria (outcomes).

Study	Outcomes						
	Outcome or exposures well defined	Appropriate sample size	Appropriate analytic methods	Variance reported	Controlled for confounding	Sufficient detail in results	Conclusions well supported
Adcock et al [6]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Alnobani et al [7]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Amodeo et al [8]	Yes	Yes	Yes	Yes	N/A[a]	Yes	Yes
Becevic et al [9]	Yes	No	Yes	Partial	Partial	Yes	No
Bettinelli et al [10]	Yes	Partial	Yes	Yes	Yes	Yes	Yes
Burke et al [11]	Partial	Partial	Partial	No	No	Partial	Yes
Duan et al [12]	Yes	Yes	Yes	No	Partial	Yes	Yes
Frazzitta et al [13]	Yes	Partial	Yes	Yes	Yes	Yes	Yes
Garingo et al [14]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Garingo et al [15]	Yes	Yes	Partial	Yes	Partial	Yes	Yes
Goldberg et al [16]	Yes	Yes	Partial	No	Partial	Yes	Yes
Holsti et al [17]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Holt et al [18]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ito et al [19]	Yes	Partial	Partial	Yes	No	Yes	Yes
Lazzara et al [20]	Yes	Partial	Yes	Yes	Partial	Yes	Yes
Marini et al [21]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Marttos et al [22]	Partial	Partial	Partial	No	Partial	Yes	Yes
McNelis et al [23]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Murray et al [24]	Partial	Yes	No	No	No	Partial	Yes
Prokazova et al [25]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Reynolds et al [26]	Yes	Partial	No	No	Partial	Yes	Yes
Rincon et al [27]	Yes	Yes	Yes	Partial	N/A	Yes	Yes
Rocca et al [28]	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rogove et al [29]	Yes	Yes	Yes	Partial	Partial	Yes	Yes
Ruiz-Del-Solar et al [30]	Partial	Yes	Partial	No	No	Yes	Yes
Shimizu et al [31]	Partial	Partial	No	No	Partial	Partial	Yes
Sucher et al [32]	Yes	Partial	No	No	N/A	Partial	Yes
Summerfield et al [33]	Yes	Yes	Partial	Yes	Partial	Yes	Yes
Vespa et al [34]	Yes	Yes	Partial	Partial	Partial	Yes	Yes
Wang et al [35]	Yes	No	N/A	Partial	Partial	Partial	Yes
Williams et al [36]	Yes	Yes	Partial	No	Partial	Yes	Yes
Ye et al [37]	Yes	Yes	Yes	Yes	Partial	Yes	Yes
Zeiler et al [38]	N/A	N/A	N/A	N/A	N/A	Yes	Yes

[a]N/A: not applicable.

Study Characteristics

The 33 studies included 4 categories of robotic presence from 10 different countries or regions: 18 (55%) from the United States, 3 (9%) from Canada, 2 (6%) from Italy, 2 (6%) from Japan, 3 (9%) from China, 1 (3%) from Chile, 1 (3%) from Switzerland, 1 (3%) from Saudi Arabia, 1 (3%) from Russia, and 1 (3%) from the United Kingdom. Of these 33 studies, 7

XSL•FO
RenderX

(21%) were RCTs, and 26 (79%) were observational studies. Patients were enrolled from 2007 to 2021. All studies were published in or translated to English. All studies involved patients in critical care settings, which included patients in the ICU, high-dependency unit, NICU, and emergency care settings where critical care had to be delivered. Characteristics of the included studies are shown in Table 3.

XSL•FO
RenderX

Table 3. Characteristics of included studies.

Study	Country or region	Study type	Setting	Population size	Use	Robot type
Adcock et al [6]	United States	Observational	ICU[a]	100 patients and 16 physicians	Telepresence	RP-7 (InTouch Health)
Alnobani et al [7]	Saudi Arabia	RCT[b]	ICU	140	Telepresence	Telemedicine Robot (Saudi Telehealth Network)
Amodeo et al [8]	Italy	Observational	NICU[c]	200 drug samples	Drug dispensing and delivery	I.V. Station (Omnicell Inc)
Becevic et al [9]	United States	Observation	ICU	5	Telepresence	RP-7 (InTouch Health)
Bettinelli et al [10]	United States	RCT	ICU	20	Telepresence	RP-7 (InTouch Health)
Burke et al [11]	United States	Observational	Emergency care	26	Telepresence	RP-7 (InTouch Health)
Duan et al [12]	China	RCT	ICU	32	Patient evaluation	MGIUS-R3 (MGI Tech Co Ltd)
Frazzitta et al [13]	Italy	RCT	ICU	40	Therapy or stroke rehabilitation	Erigo (Hocoma AG)
Garingo et al [14]	United States	Observational	ICU	46	Telepresence	RP-7 (InTouch Health)
Garingo et al [15]	United States	Observational	NICU	40	Telepresence	RP-7 (InTouch Health)
Goldberg et al [16]	United States	Observational	ICU	23 ICU bed units over a 3-year period	Telepresence	RP-7 (InTouch Health)
Holsti et al [17]	Canada	RCT	NICU	49	Therapy or stroke rehabilitation	Calmer (PCT[d] utility patient no: CA2015/051002)
Holt et al [18]	Canada	Observational	Emergency care	38	Telepresence	RP-7 (InTouch Health)
Ito et al [19]	Japan	Observational	Emergency care	9	Patient evaluation	FASTele Tele-echography robot system
Lazzara et al [20]	United States	Observational	ICU	32	Telepresence	RP-7 (InTouch Health)
Marini et al [21]	United States	Observational	ICU	28	Telepresence	RP-6 (InTouch Health)
Marttos et al [22]	United States	Observational	Emergency care	176	Telepresence	RP-7 (InTouch Health)
McNelis et al [23]	United States	Observational	ICU	14 ICU bed units over a 2-year period	Telepresence	RP-7 (InTouch Health)
Murray et al [24]	United States	Observational	ICU	69 bed units	Telepresence	RP-7 (InTouch Health)
Prokazova et al [25]	Russia	Observational	ICU	66	Therapy or stroke rehabilitation	MOTOMed LOTTO 2 (RECK-Technik)
Reynolds et al [26]	United States	Observational	ICU	22	Telepresence	RP-7 (InTouch Health)
Rincon et al [27]	United States	Observational	ICU	34 presurvey and 40 postsurvey participants	Telepresence	RP-7 (InTouch Health)
Rocca et al [28]	Switzerland	RCT	ICU	30	Therapy or stroke rehabilitation	Erigo (Hocoma AG)
Rogove et al [29]	United States	Observational	ICU and emergency care	106	Telepresence	RP-7 (InTouch Health)
Ruiz-Del-Solar et al [30]	Chile	Observational	ICU	986 visits	Telepresence	Pudu Telepresence Robot
Shimizu et al [31]	Japan	Observational	ICU	25	Telepresence	Sota (VStone Co, Ltd)
Sucher et al [32]	United States	Observational	ICU	24 patients and 26 family members	Telepresence	RP-7 (InTouch Health)

XSL•FO
RenderX

Study	Country or region	Study type	Setting	Population size	Use	Robot type
Summerfield et al [33]	United States	Observational	ICU	23 preimplementation participants, 96 postimplementation participants, and 30 for the 2-year follow-up surveys	Drug dispensing and delivery	TUG Automated Robotic Delivery System (Aethon, Inc)
Vespa et al [34]	United States	Observational	ICU	640	Telepresence	RP-7 (InTouch Health)
Wang et al [35]	China	Observational	Isolation ward	1	Patient evaluation	MGIUS-R3 (MGI Tech Co, Ltd)
Williams et al [36]	Canada	RCT	NICU	10	Therapy or stroke rehabilitation	Calmer (PCT utility patient no: CA2015/051002)
Ye et al [37]	China	Observational	Isolation ward	23	Patient evaluation	MGIUS-R3 (MGI Tech Co Ltd)
Zeiler et al [38]	United Kingdom	Observational	ICU	10	Patient evaluation	Delica EMS 9D (Shenzhen Delica Medical Equipment Co Ltd)

[a]ICU: intensive care unit.

[b]RCT: randomized controlled trial.

[c]NICU: neonatal intensive care unit.

[d]PCT: Patent Cooperation Treaty.

Benefits and Limitations of Robots

Overview

The benefits and limitations of robots can be grouped into four broad themes: (1) telepresence, (2) therapy and stroke rehabilitation, (3) patient evaluation and assessment, and (4) drug dispensing and delivery. These themes are all related to the robots' functions in various aspects of patient care in terms of monitoring, diagnostics, and treatment.

Telepresence is defined as a technology that enables a person to perform actions at a distant location as if the person were physically present at that location. Unlike other forms of remote consultation, telepresence may also include the ability to use the medical equipment of the physician, such as stethoscopes and ultrasound, allowing physicians to remotely control the robot and interact with patients and health care personnel on site. This is different from telemedicine, which involves audio or visual communication between patients and physicians in an outpatient setting and is not the focus of this study.

Therapy and stroke rehabilitation involve interventions to treat diseases, optimize functioning or reduce disability in individuals. Patient evaluation involves assessing a patient's current condition to identify health problems and plan treatment. Finally, drug dispensing and delivery involve the process of preparing and providing medicine to a patient based on a health care provider's prescription.

Theme 1: Telepresence

Approximately 64% (21/33) of studies identified 5 different telepresence robots. RP-7 (InTouch Health) [6,9-11,14-16,18,20-24,26,27,29,32,34] was the main robotic telepresence system used in 55% (18/33) of studies (RP-6 was used in one). RP-7 has a bidirectional audio and video communication system that displays real time video and camera systems. Devices such as electronic stethoscopes, otoscopes, pulse oximeters, and ultrasound probes can be connected to the expansion bay of a robot to transmit medical data. The robotic system can be remotely controlled and monitored by physicians. RP-7 can also be linked to and automatically acquire information from hospital-based electronic data systems.

Sota (Vstone Co Ltd) [31] is a bedside AI-enhanced robot capable of alerting physicians about anomalies in biological information. Such information can be derived in real time from bedside monitors or existing electronic health records. The Sota robot alerts physicians through voice warning systems coupled with alarms. In addition to the alert function, it can function as a social robot by responding to simple voice commands. In contrast to Sota, Pudu [7] is a social robot designed specifically to provide telepresence and communication services and deliver emotional and mental care to isolated patients with COVID-19. It works by using an assistive teleoperation mode, allowing for remote control of the robot's movements using an Xbox (Microsoft) controller joystick. The robot comprises smooth surfaces and fulfills health requirements, where it can be sanitized in a safe and efficient way. Two unnamed telepresence robots were used in a study in Saudi Arabia [30]. These 2 robots had similar functions and equipment to RP-7.

Patients benefited from telepresence because of the reduced response time [11,14,18,21,24,29,34] by as much as 95.8%, allowing earlier intervention, higher patient survivability, and lower mortality rates [16,21]. Unlike more traditional methods, the physician was able to have a realistic physical presence and interact directly with ICU staff and patients at the bedside [14]. Mortality and complication rates could be reduced by 25% [21] to 59% [16], especially at night when there were often staff shortages [16]. This was especially pertinent in time-sensitive settings such as trauma with a short time window to intervention [34] and rural hospitals with poor access to specialist physicians [22,24,26]. Overall, telepresence allowed care to be provided

XSL•FO
RenderX

in a timely manner regardless of the location of the physician or the time of day.

When compared with care delivered by traditional methods, the studies that measured rates of ICU admission and the average length of stay consistently showed a decrease in length of stay compared with both conventional rounding and telephone rounding, ranging from 6.25% [24] to 33% [16], and an increase in appropriate ICU admissions rates [16,21,23,24,34]. With prompt response time and closer monitoring, the rates of developing significant complications were lower. Any emergencies or acute changes were tended to before significant health repercussions developed [6].

The usual standards of care and assessment were not compromised when the robots were used. Approximately 12% (4/33) of studies mentioned that the RP-7 robot was able to perform a good range of tasks, including physical examinations, with a similar level of accuracy and precision compared with traditional methods of care, allowing the physician to come to an accurate clinical conclusion [6,11,14,21].

For hospitals, there was a financial benefit from direct cost savings as robotic presence reduced the need to employ full-time staff for ward rounds during off-peak hours [16,23,34]. There were also cost savings from faster patient turnover and the lowered external transfer rate of rural hospitals [28]. By reducing the number of external transfers, the number of unnecessary admissions to hospitals was reduced. In total, the financial benefits were as much as US $1.1 million per year [34].

Robots were well-received by patients, family members, and staff [7,9,10,15,20-23,26,27,31]. Despite a telepresence robot providing remote physician presence, patients did not perceive the physician to be caring less or compromising the quality of care [7,15,29,32]. Staff had an overall positive perception of telepresence robots, including in areas such as usability [7], acceptability [22], efficiency, communication [9], and decreased noise or traffic in ICUs during the morning rounds [32]. For example, in a study by Alnobani et al [7], 71.5% of staff felt that the robot saved time, and 77.2% of staff felt that it improved clinical diagnosis.

Robots also played a role in the education and mentoring of staff [21,26]. Staff education included mentoring nurses, discussing admission and discharge issues, and facilitating compliance with treatment protocols. Expert opinions from nurses and physicians were more accessible for direct guidance of resuscitation efforts, even in remote areas [26]. Interactivity and 2-way communication were preserved during the teaching that occurred during remote rounding [21]. In addition, hospital psychologists used the Pudu robot [30] to provide remote emotional and mental care in the COVID-19 ward. All patients who received such psychological care via Pudu showed positive attitudes and emotions. Patients and family members were satisfied with how Pudu enabled their interactions to be extended and uninterrupted, providing them with good emotional support.

Limitations of telepresence included discrepancies between on-site and off-site evaluations, although these could be attributed to subjective differences [14]. One of the studies reported limitations in determining abdominal distension and capillary refill time and using an electronic stethoscope for heart, breath, and bowel sounds [14]. However, the study also mentioned that these discrepancies were present between 2 bedside neonatologists, thus rendering it possible that these differences in findings were inherently subjective. Another study reported limitations in accurate assessments using the Mayo Full Outline of Unresponsiveness scale, particularly for brainstem and pupillary responses [6]. However, the study also reported that the Glasgow Coma Scale was a good alternative that was accurately assessed using the telepresence robot.

Medicolegal challenges existed, such as a lack of established protocols causing regulatory barriers in terms of obtaining credentialing and malpractice liability [7,29], as well as financial barriers in terms of patient billing and difficulty obtaining reimbursement [29]. In addition, hidden costs for maintenance and electricity, licensing, technical issues, and space constraints acted as barriers to use [29].

Although many studies mentioned a reduction in face-to-face response time, 6% (2/33) of studies reported an *increase* in time spent on patient encounters, attributed to the time taken to operate and maneuver the robot, as well as to resolve technical issues such as internet connectivity problems [15,23]. Similar technological limitations of internet connectivity and maneuvering difficulty were also reported in another study [14]. Fortunately, most incidents of poor connectivity were promptly overcome within 5 minutes. Additional technological difficulties included poor audio quality because of transmission of ambient noise and poor angle of visibility when attempting to view the thoracoabdominal area [22].

In terms of staff perception, some concerns were raised with regard to the impact of robot use. These were in the areas of threat to staff job security and additional responsibilities [7]. The staff also raised some issues with regard to patient confidentiality, patient privacy, and legal liability. Nonetheless, although these concerns existed, there was general acceptance and approval of telepresence technology among the staff surveyed [7]. The benefits and limitations in the field of telepresence are summarized in Textbox 1.

XSL•FO
RenderX

Textbox 1. Theme 1: robotic telepresence.

Robot examples

- RP-7 (InTouch Health): 18 papers [6,9-11,14-16,18,20-24,26,27,29,32,34]
- SotaTM (VStone Co Ltd): 1 paper [32]
- 2 unnamed telerobots: 1 paper [31]
- Pudu Telepresence Robot: 1 paper [30]

Benefits

- Patient survival and patient mortality rate [16,21]
  - 59% lower mortality rate [16]
  - 25% decrease in mortality from robotic telerounding vs conventional rounding [21]: 12% (5/42) vs 16% (6/37); *P*=.75
- Provides superior care to alternatives [18,23]
  - Higher average number of therapeutic interventions vs telephone rounding [23]: 5.3 (SD 1.7) vs 1.3 (SD 1.4); *P*<.01
  - Less overnight calls and less unexpected events vs telephone rounding [23]: 0.1 (SD 0.2) vs 1.3 (SD 0.5); *P*<.05
  - Reduced external transfer rate by 63%, allowing patients to be effectively treated in local clinics [18] and receive specialist care closer to home and earlier stabilization
  - Patient care time can be lengthened to allow for extended interaction with family members for those under isolated care without risk of contagion exposure [30]
- Reduction in face-to-face response time, leading to earlier intervention and access to specialists [11,14,18,21,24,29,34]
  - Response latency in robotic telepresence vs conventional care [34]
    - To routine and urgent pages: 9.2 (SD 9.3) minutes vs 218 (SD 186) minutes; *P*<.001
    - To brain ischemia: 7.9 (SD 2.8) minutes vs 152 (SD 85) minutes; *P*<.001
    - To elevated intracranial pressure: 11 (SD 14) minutes vs 108 (SD 55) minutes; *P*<.001
- Decreased intensive care unit length of stay [16,21,23,24,34]
  - Length of stay in intensive care unit decreased; response latency in robotic telepresence vs conventional rounding:
    - 7.5 (SD 8.8) days vs 8 (SD 8.3) days [34]
    - 33% reduction [16]
    - 2.5 days vs 3.3 days [24]
    - 5 (SD 2) days vs 6 (SD 3) days; *P*=.57 [21]
  - Length of stay in intensive care unit decreased; response latency in robotic telepresence vs telephone rounds [23]: 4.8 (SD 2.6) days vs 5.6 (SD 2.2) days; *P*<.05
  - Length of stay in hospital decreased; response latency in robotic telepresence vs telephone rounds [23]: 10.2 (SD 4.3) days vs 12.3 (SD 4.4) days; *P*<.05
- Financial benefit: decreased cost, increased revenue, lower start-up costs or flexibility, and no need to employ full-time staff such as in the central monitoring model [16,18,34]
  - 29% lower adjusted mean direct cost estimated per case [16]
  - US $1.1 million cost savings over 1 year [34]
  - CAD $360,000 (US $285,420) savings over the study period [18]
    - Cost of round trip, cost of hospital stay, and miscellaneous costs such as family transport and accommodation
- Does not compromise on usual standard of care and assessment consistency between bedside and remote examination [6,11,14,21,30]
  - Bedside vs remote examination [6]
    - Mean Glasgow Coma Scale: 7.5 (SD 3.67) vs 7.23 (SD 3.85), difference 0.25 (SD 0.10); *P*=.01; however, the difference is not clinically significant; Pearson correlation coefficient=0.97
    -

XSL•FO
RenderX

Mean Full Outline of Unresponsiveness: 9.63 (SD 4.76) vs 9.21 (SD 4.74), difference 0.40 (SD 2.00); *P*=.05; Pearson correlation coefficient=0.91

- Agreements in most physical examination assessments between both on-site and off-site neonatologists [14]

- Education benefits [21,26,34]
  - Educational experience of medical students, physician assistants, and surgical residents not affected by response latency in robotic telepresence [21]; average Likert score:
    - Surgical residents: 4.5 (SD 0.2); *P*>.05
    - Medical students: 3.9 (SD 0.4); *P*>.05
    - Physician assistants: 4.4 (SD 0.4); *P*>.05
  - 87% felt that it improved nursing education [26]
- Positive staff perception: usability, acceptability, efficiency, communication, and decreased noise or traffic [7,9,10,15,20-23,26,27,30,31]
  - Positive health care worker attitude toward telepresence [7]
    - Increasing communication and collaboration among providers: 4.01/5 (SD 0.800)
    - Improve clinical decisions: 3.91/5 (SD 0.877)
    - Provide access to specialized second opinion consultation: 4.19/5 (SD 0.774)
    - Facilitates diagnosis and treatment: 3.87 (SD 0.847)
  - Collaboration and Satisfaction About Care Decisions survey increased [10]
    - RP-7 vs baseline: 51.3 vs 43.0; *P*=.01
    - Robot rounds vs telephone rounds: 51.3 vs 50.5; *P*=.30
  - Higher user satisfaction vs telephone rounds [23]: 7.7 (SD 2.3) vs 5.6 (SD 2.1); *P*<.01
  - Night nurses' perceptions [27]:
    - Intensive care unit physicians sufficiently available: 6%-20%; difference in proportions 14%; *P*=.008
    - Present during acute emergencies: 44%-65%; difference in proportions 21%; *P*=.007
  - SotaTM: alerts issued by the robot to warn of detected anomalies perceived to be more effective than the current desktop-based system [31]
- Patient, family perception, or satisfaction [15,26,32]
  - 100% viewed it as valuable in improving family and patient satisfaction [26]
  - 100% of parents felt comfortable talking to off-site neonatologists on a mobile robot [15]
  - 84% believed that care was better as the robot was used [32]
- Alleviate future staffing shortages
  - Allow for redistribution, easing the overcapacity issues that strain tertiary care centers [18]

Limitations

- Lack of established protocols [7,29]
  - Hindered by regulatory barriers of licensing, credentialing, and malpractice protection [29]
  - Increases legal liability challenges [7]: 2.66/5 (SD 0.784)
- Hidden costs [29]
  - Finance barriers of miscellaneous costs, billing, and reimbursement issues [29]
- Discrepancies between on-site and off-site evaluations for physical findings [7,14]
  - Poor agreements on physical examination parameters (breath, heart and bowel sounds, and capillary refill time) [14], although they also occurred regardless of response latency in robotic telepresence use between 2 on-site physicians

XSL•FO
RenderX

- Decreased efficiency and longer time spent on patient encounters
  - Time spent [15] off-site vs on-site neonatologist: 8 (IQR 7-10.5) minutes vs 5 (IQR 5-6) minutes; *P*=.002; difference because of time needed to operate and maneuver robot or slower or dropped internet connection
  - Longer rounding time [23] response latency in robotic telepresence vs telephone: 33.2 (SD 15.4) minutes vs 18.3 (SD 12.7) minutes; *P*<.05
- User-dependent experience required training [7,21]
- Technological limitations
  - Difficulties maintaining internet connection in 23% encounters; 93% reconnected in <5 minutes [14]
  - Average of 2.1 (SD 1.2) interruptions per session because of wireless signal loss [23]
- Ethical challenges [7]
  - Threatens patient's confidentiality: 2.96/5 (SD 0.955)
  - Raises privacy concerns: 3.12/5 (SD 0.956)
- Poor staff perception [7,9,20,21,26,31]
  - 50% of physicians did not think physician quality of life improved [26]
  - Did not meet nurses' expectations [21]; Likert score of 3.5 (SD 1.0)
  - Threatens staff position [7]: 3.09/5 (SD 0.925)
  - Increases staff workload [7]: 3.09/5 (SD 0.925)
  - Creates new responsibilities for staff [7]: 2.74/5 (SD 0.940)
  - Only 20% of nursing respondents were satisfied with the quality of technology of Sota Robot [31]

Theme 2: Therapy and Stroke Rehabilitation

Approximately 15% (5/33) of studies identified 3 different robots (Table 4) that provided various forms of therapy or rehabilitation in the intensive care context. They played a role in enhancing and optimizing the process of patient recovery. Of the 3 robots we identified, 2 (67%) were targeted toward early functional rehabilitation for patients with stroke [13,25,28], and 1 (33%) was a robot specially designed for the care of neonates in the NICU [17,36]. Physiological parameters were measured to evaluate the effects of the 3 robots, which have been shown to be beneficial overall.

MOTOmed LOTTO 2 (RECK-Technik) [25] is a robotic movement therapy device that enables leg mobilization in a supine position, allowing for passive, active, or assisted mobilization for patients on prolonged bed rest. Early rehabilitation of patients of stroke has been shown to lead to better functional outcomes in patients with acute ischemic stroke [39]. MOTOmed [25] achieved better outcomes than a standard care protocol in terms of recovery of neurological function. MOTOmed also achieved a lower incidence of severe multicomponent multiple organ dysfunction (14% vs 41%; *P*<.05; intervention vs control) and pulmonary embolism (12% vs 33%; *P*<.05; intervention vs control). In patients with neurological pathologies, MOTOMed stimulated the sympathetic system, which helped recovery by preventing polyneuropathy and improving awareness of disorders of consciousness. However, MOTOMed should be used with caution in patients with subarachnoid hemorrhage, as catecholamine overproduction as a stress response was associated with complications such as the increased risk of vasospasm [28].

Erigo (Hocoma AG) [13,28] is a robot that combines a tilt table with a leg movement system, allowing for progressive and customizable verticalization of patients with acquired brain injury. The gradual mobilization in Erigo overcame an important limitation to early mobilization, which was orthostatic intolerance [13,28]. Orthostatic hypotension with compensatory sympathetic catecholamine production was reduced most significantly with Erigo compared with other forms of early mobilization, namely conventional in-bed physiotherapy and MOTOMed [28]. Therefore, it could be safely used in patients with subarachnoid hemorrhage. Compared with in-bed physiotherapy, Erigo produced statistically significant, higher improvements in the Coma Recovery Scale (17.0 vs 5.0; *P*=.03; intervention vs control) and Disability Rating Scale (−20.0 vs −6.0; *P*=.04; intervention vs control) [13]. It also produced nonstatistically significant improvements in the Glasgow Coma Scale and levels of cognitive functioning. However, a longer ICU stay was required to complete the verticalization protocol before transfer to a neurological rehabilitation unit [13].

Calmer [17,36] is a robot used in the NICU, which is designed to reduce pain in preterm infants subjected to multiple painful procedures. Calmer simulates skin-to-skin holding via touch, breathing motions, and sound stimulation. Calmer's artificial skin-like surface and vertical movement mimic breathing motion and heartbeat sound to match those of infants' mothers. Compared with the standard care of facilitated tucking, Calmer reduced preterm infant pain reactivity. Approximately 6% (2/33) of studies consistently showed that infants had greater *para*sympathetic activation and hence greater physiological stress *reduction* during painful procedures such as blood taking [36]. Calmer was a safe, ergonomic, and cheaper alternative to

XSL•FO
RenderX

the manpower-intensive facilitated tucking. Research is ongoing to incorporate Calmer into incubators, which would potentially allow for cost savings of as much as US $380,000 per year in a 60-bed NICU [17]. The benefits and limitations of the 3 abovementioned robots are summarized in Table 4.

XSL•FO
RenderX

Table 4. Theme 2: therapy or stroke rehabilitation.

Robot examples	Benefits	Limitations

Calmer: 2 papers [17,36]	• Efficacy in reducing infant pain
	  • Increases HF[a,b] component (parasympathetic activity) of HRV[c] (Hz/ms2) [36], Calmer vs standard FT[d] group:
	  • Baseline (before procedure): 36.0 (23.7-73.2) vs 3.6 (3.1-9.1)
	  • Poke (during painful procedure): 2.2 (1.1-3.0) vs 0.4 (0.3-7.2)
	  • Recovery (post procedure): 6.8 (1.7-21.1) vs 5.2 (4.1-12.8)
	  • Difference BIIP[e] score in peak pain phases, Calmer vs FT [17]: 4.0 (SD 2.7) vs 3.2 (SD 2.7; 95% CI −0.45 to 2.72)
	• Cost savings
	  • US $380,000 per year in 60-bed NICU[f] [17]
	• No safety issues with short-term use
		Nil mentioned

MOTOmed LOTTO 2 (RECK-Technik): 1 paper [25]	• Safe for early rehabilitation of patients of stroke who are critically ill
	• Better outcomes in stroke rehabilitation (day 21 after stroke), intervention vs control group:
	  • Neurological outcomes improved
	  • GCS[g]: 15 (14-15) vs 15 (15-15); *P*=.32
	  • NIHSS[h]: 11 (8-25) vs 15 (12-19); *P*>.05
	  • APACHE[i] 2: 6 (3-14) vs 9 (6-12); *P*>.05
	  • Complications
	  • Incidence of MOD[j]: 60% vs 67%; *P*>.05
	  • Incidence of severe MOD: 14% vs 41%; *P*<.05
	  • MOD scale 0 (0-1) vs 1 (0-2); *P*>.05
	  • Incidence of PE[k]: 12% vs 33%; *P*<.05
	  • Incidence of death from PE: 0 vs 1/3
	• Mortality rate decreased, intervention vs control group: 12% vs 39%; *P*<.05
		• No significant changes in DVT[l] incidence, intervention vs control group:
		  • DVT incidence 58% vs 45%; *P*>.05

Erigo (Hocoma AG): 2 papers [13,28]	• Better clinical outcomes—greater difference in neurological scoring systems
	  • Difference in values at ICU[m] admission and at rehabilitation discharge [13], intervention vs control:
	  • No orthostatic intolerance occurred
	  • DRS[n]: −20.0 (−22.0 to −4.5) vs −6.0 (−12.7 to −2.0); *P*=.04
	  • CRSr[o]: 17.0 (5.1-18.8) vs 5.0 (2.4-11.0); *P*=.03
	  • GCS: 7.0 (3.2-10.0) vs 4.5 (3.0-6.5); *P*=.08
	  • LCF[p]: 4.0 (1.0-5.0) vs 2.5 (1.0-4.0); *P*=.14
	• No increase in catecholamine production [28]
		• Longer LoS[q] in ICU [13], intervention vs control group: 38.8 (SD 15.7) days vs 25.1 (SD 11.2) days; *P*=.01
		  • To complete stepping verticalization protocol before being moved to the neurological rehabilitation unit

[a]HF: high frequency.

[b]Indicates parasympathetic activity: decreased HF=stress; increased HF=calmness or stress recovery.

[c]HRV: heart rate variability.

[d]FT: facilitated tucking.

[e]BIIP: Behavioural Indicators of Infant Pain.

[f]NICU: neonatal intensive care unit.

[g]GCS: Glasgow Coma Scale.

[h]NIHSS: National Institutes of Health Stroke Scale.

[i]APACHE: Acute Physiology and Clinical Health Evaluation.

[j]MOD: multiorgan dysfunction.

XSL•FO
RenderX

[k]PE: pulmonary embolism.

[l]DVT: deep vein thrombosis.

[m]ICU: intensive care unit.

[n]DRS: Disability Rating Scale.

[o]CRSr: Coma Recovery Scale.

[p]LCF: levels of cognitive functioning.

[q]LoS: length of stay.

Theme 3: Patient Evaluation and Assessment

Approximately 15% (5/33) of studies identified 3 different robots (Table 5). These robots were used to evaluate various parameters of patients, including patient monitoring in a critical care setting and ensuring quality evaluation from a remote location. Robots also used ultrasound systems to enhance their evaluation capability [12,19,35,37,38].

FASTele [19] is a wearable, portable, attachable tele-echography robot system for focused assessment with sonography for trauma (FAST) scans. FASTele [19] was able to produce sharp ultrasound images of all FAST areas, even under maximum vehicle acceleration, in all axial directions, and under various body motion conditions, applicable to a range of body types. However, a longer time was required to perform FAST, especially in patients who were overweight, as it required attaching a corset to each FAST area. Patients were also at risk of injury during the attachment of the robot system.

MGIUS-R3 (MGI Tech Co Ltd) [12,35,37] is a 5G-powered, remote, robot-assisted teleultrasound diagnostic system. It combines a robotic arm, an ultrasound imaging system, and audio-visual communication for teleoperation. Application of MGIUS-R3 in cardiopulmonary assessment achieved image quality acquisition, labeling, and analysis equivalent to that of traditional ultrasound, enabling accurate diagnosis [12,37]. No complications or delays were noted during the image acquisition process. It performed satisfactorily even at remote distances of 700 km [35]. Overall, there was a higher level of safety because of the reduction in infection risk for patients and physicians. The patient would not be exposed to cross-infection during transport to the radiographer's room in the hospital, and the physician would not be exposed to a patient with an infectious disease [12]. However, the robotic arm faced difficulties in reaching some body parts [12,37]. In addition, the ultrasound frequency was limited as the robot had only 1 convex array probe, thus affecting the quality of cardiac images [12,37].

Delica EMS 9D (Shenzhen Delica Medical Equipment Co Ltd) [38] is a portable transcranial Doppler (TCD) system for simultaneous bilateral middle cerebral artery blood flow velocity recording. It comprises Doppler ultrasound probes attached to a robotic drive supported by a headband frame. The robot performs the automated functions of scan, search, direction, and track. Compared with standard TCD systems, Delica achieved improved image-capturing capabilities without interruption or the need for manual adjustments [38]. Delica also reduced the risk of disrupting other in situ monitoring, making it highly applicable to the critical care setting. Overall, this led to time saving and increased efficiency. However, this device has some technological limitations. The only available signal recording frequency was 100 Hz, and it could not perform heart rate variability analyses requiring frequencies of ≥200 Hz. A possible safety concern involved increased intracranial pressure in one patient, which was resolved with headband readjustment. The benefits and limitations of the 3 abovementioned robots are summarized in Table 5.

XSL•FO
RenderX

Table 5. Theme 3: patient evaluation.

Robot examples	Benefits	Limitations
FASTele: 1 paper [19]	• Extracted echo images met and exceeded the defined FAST[a] criteria • Brightness gradient of echo images vs values required by the physician: 4.7 (SD 10.4) vs 3.9 (SD 9.8) • FAST performance achieved with vehicle motions: at maximal acceleration in all axial directions and body motion conditions • Constant pressure to hold the probe is not required	• Likelihood of longer time to perform FAST: requires attaching a corset to each FAST area and may cause possible injury to patients • Prolonged wrapping time in patients who are overweight • System to be improved for medical physicians to operate it easily

MGIUS-R3 (MGI Tech Co Ltd): 3 papers [12,35,37]	• Clear images: image quality score 4.73 (high quality) [12]
	• Comparable diagnostic results to bedside examination [12,37]
	• Safety [12,37]
	  • Able to complete an assessment successfully as per established examination protocol [12,37]
	  • No need to transport patients who are clinically ill for assessment and minimizes radiographer and hospital exposure to COVID-19 and other infectious diseases [12,37]
	  • Able to be used in isolation wards [37]
	  • Multiple protection measures [37]
	  • Simultaneous start prompts
	  • Emergency stop button
	  • Speed limit settings on the robotic arm
	• Faster [12,37]
	  • No delay in scanning, 10-20 minutes per examination [37]
	  • 5G network system: ensures real time US[b] image; detailed physician-patient communication, 20 times better transmission rate; delay reduced by a factor of 10, allowing high-definition and accurate video transmission [12,37]
	• Able to perform from a remote distance of 700 km away [35]
		• Difficulty of the robotic arm in reaching some body parts, especially in patients who are critically ill [37] and on the patient's side [12]
		  • Required mobilization of intubated COVID-19 patient for AP[c] and lateral thoracic views [35]
		• Only one convex array probe—frequency limitation and unable to scan heart [12,37]
		• 15.6% inconsistent results between robot-assisted teleultrasound and bedside ultrasound [12]
		• Difficulty in 3D space perception, requiring practice and familiarization [12]

Delica EMS 9D robotic TCD[d] (Shenzhen Delica Medical Equipment Co Ltd): 1 paper [38]	• Improved image-capturing capability vs standard TCD systems
	  • Continuous, uninterrupted recording for 4 hours
	  • Better image quality
	• Flow velocity signals are accurately captured even in the presence of other in situ multimodal monitoring devices
	  • Allows multimonitoring in moderate to severe TBI[e] patients
	  • Reduces risk of disruption of monitoring from repeated loosening and manipulation of other devices
	• Increased efficiency from time saved in manual adjustment of the probe, which is crucial in patients who are critically ill
	• Increased patient comfort and fast turnover with easy cleaning of the device
		• Scan and track functions are less functional
		• Limitation in available signal recording frequency (100 Hz only)
		• Potential complications of raised ICP[f]

[a]FAST: focused assessment with sonography for trauma.

[b]US: ultrasound.

[c]AP: anterior posterior.

[d]TCD: transcranial Doppler.

[e]TBI: traumatic brain injury.

XSL•FO
RenderX

[f]ICP: intracranial pressure.

Theme 4: Drug Dispensing and Delivery

Approximately 6% (2/33) of studies identified 2 different robots (Table 6). Both robots were involved in drug dispensing or delivery, and both showed time reduction, cost savings, and increased precision in drug preparation.

The TUG Automated Robotic Delivery System (Aethon Inc) [33] is a robot affixed to a medication delivery cart controlled by pharmacy staff. When a medication delivery was planned, a pharmacy staff member summoned the robot, inputted the desired sequence of deliveries, and loaded the medications onto the robot. The robot then traveled to the desired locations where the nurses unloaded it. The robots delivered most medications except for *stat* medications meant for immediate administration and controlled drugs. The TUG robot reduced the mean pharmacy cycle time from order receipt to order exit by 29.6% [33]. The technician delivery time decreased by 7.2 hours, and the saved time was used in handling other pharmacy tasks, leading to significant cost savings of an estimated US $14,100 yearly. It was well-received by nurses and pharmacists in terms of reliability and performance. However, nurses were dissatisfied that they now had to sort and store medications, which had been previously performed by technicians. There was also a downtime of robots because of infrastructure and robot-related problems such as power supply and cart issues.

I.V. Station (Omnicell Inc) [8] is a fully automated robot that prepares sterile injectable drugs. It performs all stages of preparation, from reconstitution to dilution and final preparation. I.V. Station achieved increased precision in drug preparation compared with manual preparation [8]. Patient adverse effects from overdosing and loss of drug efficacy from underdosing were reduced. In addition, there were fewer potentially harmful staff events. Precision is especially crucial for preterm neonates who require complex therapy and are at high risk of fatal medication errors [8]. A decrease in the cost and mean preparation time by as much as 8% and 2 hours 57 minutes, respectively, was achieved during the preparation of greater quantities. Time savings enabled a focus on other aspects of care, including engaging and educating families. However, when preparing smaller quantities, the robot was more expensive and slower than manual preparation. In addition, mechanical or software failure events affected the workflow and caused medication wastage. The benefits and limitations of the 2 abovementioned robots are summarized in Table 6.

Table 6. Theme 4: drug dispensing and delivery.

Robot examples	Benefits	Limitations

TUG Automated Robotic Delivery System (Aethon Inc): 1 paper [33]	• Increased efficiency of medication delivery before implementation vs 2 years after implementation
	  • Mean total mean pharmacy cycle time (order receipt to order exit): 73.9 (SD 2.21) minutes vs 52 (SD 28.6) minutes
	  • Mean time for label printing, 13.1 (SD 3.9) minutes vs 7.4 (SD 4.1) minutes
	  • Mean idle time for medication delivery: 27.3 (SD 8.2) minutes vs 15.3 (SD 8.4) minutes
	• Time and cost savings
	  • 7.2 hours of technician time saved
	  • Projected annual cost savings: US $14,100
	• Positive nurse perceptions:
	  • Perception before implementation vs post implementation:
	  • General satisfaction increased; $P<.02$
	• Robot reliability increased; $P<.01$
		• Limited benefit in timeliness and perceived quality of delivery service
		• Decreased efficiency in nondelivery aspects—nurses have additional duty to sort and store delivered medications.
		• Low robot reliability perceived by technicians that improved at 2-year follow-up

I.V. Station (Omnicell Inc): 1 paper [8]	• Better clinical outcomes
	• Increased precision in drug preparation vs manual preparation: accuracy within 5% to –5%
	• Improved safety for both patient and staff
	• Increased efficiency during the preparation of higher dose quantities
	  • Range: time savings of 16 seconds (acyclovir) to 2 hours 57 minutes (teicoplanin)
	• Reduced costs during the preparation of higher dose quantities
	  • Range: 8% (ampicillin) to 66% (teicoplanin)
		• Mechanical or software failure events
		• Decreased efficiency during the preparation of lower dose quantities
		• Increased costs during the preparation of lower dose quantities
		• Hidden costs (not included in cost calculations)
		  • Electricity
		  • Machine maintenance
		  • Days of downtime because of machine failure
		• However, the inactivity rate was low at 2.5% (9.5/365 days)

XSL•FO
RenderX

Discussion

Benefits of Robots

Our review demonstrates the numerous beneficial capabilities of robots. We found that the greatest application of robots in critical care was in telepresence, and the most studied telepresence robot was RP-7. Overall, the evidence showed that robots were beneficial and well-received and delivered significant patient, staff, and hospital benefits. The abovementioned robots covered various aspects of ICU care. Some were used during acute settings, such as telepresence robots for urgent consultations or for patient evaluation. Meanwhile, the robots that focused on rehabilitation or drug dispensing were more directed toward general functioning and processes in the ICU.

In terms of efficiency, robots in the areas of telepresence, patient evaluation, and drug dispensing and delivery were able to provide time savings. In the critical care setting, this was especially important, as face-to-face response time could be reduced, allowing patients to have faster access to specialists.

Similarly, there were cost savings in the applications of telepresence, therapy or rehabilitation, and drug dispensing or delivery. Although the amount saved varied across different studies, with the highest being US $1.16 million reported by Vespa et al [34], all studies agreed that cost savings were beneficial to hospitals.

Robots could outperform current care standards and supplement human efforts in the fields of telepresence, therapy, and patient evaluation. For example, Delica TCD [38] allowed for improved Doppler image capturing that a normal TCD could not achieve with manual effort. With Erigo [13,28], concurrent verticalization with stepping eliminated orthostatic hypotension, which previously prevented early mobilization post acquired brain injury. This enabled improved care for patients with a subsequent reduction in mortality rate.

The workload of physicians could also be alleviated using robots. RP-7 [6,9-11,14-16,18,20-24,26,27,29,32,34] supplemented rounding and was used during off-peak hours, reducing the need for physical physician presence during graveyard shifts. This is particularly relevant during the current COVID-19 pandemic, where physicians must grapple with a heavy workload [40]. Physicians could then focus on more holistic patient care, including psychological and social aspects.

A benefit mentioned across all themes was safety. Generally, papers in each theme agreed that robots were able to either meet current safety standards by providing diagnoses comparable with those of existing standardized methods or further reduce risks, for example, by improving the precision of medication preparation [8]. In more recent papers published in 2020 and 2021, a consistent theme was that robots could allow medical professionals to maintain social distancing while still effectively treating patients. This prevented exposure to pathogens and also reduced the use of disposable personal protective equipment.

Many believe that robots are unequipped to handle soft skills instrumental in health care. Although robots cannot counsel a patient or console a distressed family member, they can nonetheless emulate the human touch in their own unique ways. For example, Calmer [17,36] sought to mimic human touch without the intention to *replace the parent*. The technology is a step in the right direction, and the comfort that the robot brings to infants could potentially be extended to the care of adult patients who are vulnerable and critically ill as well.

Limitations of Robots

Although robots could help reduce the workload in some areas, they could lead to both human unemployment and overreliance on robots. Although robots cannot fully replace physicians, they can and already have replaced some manpower in the health care sector. When surveyed, staff in the ICU felt that their jobs were moderately threatened [7].

Another concern was the possibility of hacking. Some robots such as RP-7 and MGIUS-R3 relied on Wi-Fi or 5G and thus were susceptible to security issues and data breaches. If robots were to break down or encounter technological issues, systems must be in place to immediately recognize and mitigate these issues, given that time is always of the essence in health care. Otherwise, cybersecurity breakdowns would lead to workflow disruptions, loss of patient privacy, and significant medicolegal repercussions [7].

Robot use may translate to increased costs for patients because of the cost of robots, licensing, installation, maintenance, and repairs. In addition, because of the current lack of legislation regarding billing for services rendered by robots, hospitals may excessively charge for robotic use. Goldberg et al [16] reported that although the mean cost estimates per case decreased by 29%, the billing charges instead increased by 70% [16]. Overall, this could mean that although costs for hospitals decreased, costs for patients ultimately increased, which also reflects a mismatch in expected outcomes, possibly because of a lack of existing price controls.

We must also recognize that telepresence implementation may be more suited to hospitals that already have an effective ICU staffing infrastructure. Although the ideal aim of telepresence is to relieve the workload of ICUs with scarce resources, it may potentially create a paradoxical imbalance in resource allocation, where staff from underresourced ICUs are drawn to larger, more established ICUs that can sustain telepresence.

Although robots complement and aid in workload, leading to generally positive perceptions, some robots were less enthusiastically received. This could be attributed to differences in the ease of use of the robot, the context of their application, and baseline perceptions. One of the studies mentioned that nurses still believed that the physical presence of intensivists was preferable and necessary. In any case, the role of telepresence is not to completely replace physical physician presence but to supplement staffing during off-peak hours, ensuring safe coverage.

As robotic intervention becomes more prevalent and integrated into health care, this necessitates a conversation around developing an ethical and legal framework with regard to accountability. With exponential digital growth over the past century, we will certainly continue to see an increasing overlap

XSL•FO
RenderX

between the physical and digital worlds. It is imperative to form strict lines of accountability—shall it lie with the physician who used it, or is the robot's developer and manufacturer who should be held accountable for any errors?

Further Applications

Among the excluded papers, promising potential applications of robots were shown, as elaborated in the following sections.

The COVID-19 Pandemic

In the past 2 years, research has been greatly focused on the COVID-19 pandemic. As many papers have yet to have formal trials on patients, they were excluded based on our criteria. However, they demonstrate highly applicable uses in the critical care setting. Overall, 7 COVID-19-related papers echoed similar themes to those papers included in our systematic review. In addition, by enabling remote disinfection or control of equipment, robots reduced the exposure of medical staff to pathogens and lowered the use of personal protective equipment.

Approximately 29% (2/7) of papers described the use of UV-C disinfectant robots within the ICU [41,42]. Choi et al [41] described a UV light-emitting diode robot (UVER-SR1, UVER Co) [41] with a freely rotating arm. It could successfully disinfect ICU rooms. Another mobile UV-C robot (ASSUM, Assum Tech) [42] similarly demonstrated a 99.91% reduction in the SARS-CoV-2 load within a few minutes. Overall, the 2 robots worked to reduce the exposure of cleaning and health care personnel to contaminated surfaces.

Approximately 57% (4/7) of papers described the use of robotics to reduce the need for health care staff to physically enter patient rooms. Sawyer (Rethink Robotics GmbH) [43], a 7-axis robot with flexible joints and a camera, could successfully perform a variety of COVID-19 health care tasks: intravenous pump device continuation, ventilator knob adjustment, ICU monitor silencing, oxygen knob adjustment, and call button deactivation. Vagvolgyi et al [44] demonstrated the use of a telerobotic cartesian system to allow the adjustment of ventilator settings from outside the ICU. Similarly, this was feasible in a simulated ICU environment and specifically saved 59.8% of the time (a decrease from 271 seconds to 109seconds). A 4-Degrees-Of-Freedom [45] robot was also able to interact with the touchscreen instrument panel of dialysis machines, achieving fast and simple control of the machine in the context of emergency dialysis. Finally, in a newly published study, a team from the Massachusetts Institute of Technology developed Emergency-Vent [46], a robotic gripper that automated the task of manually squeezing a resuscitator bag. The robot was able to customize ventilator settings within each cycle of breathing, including tidal volume, respiratory rate, inspiration-expiration time ratio, positive end–expiratory pressure, and assist control trigger threshold. It successfully ventilated a porcine model and performed comparably to that of an experienced anesthesiologist manually pumping the resuscitator bag. Both ease and cost of use were low. Such robotic ventilator technology is especially relevant in the context of the COVID-19 pandemic, as manpower and ventilator shortages lead to the need for cheap ventilator alternatives.

Long-term bed immobilization limits the recovery of patients with COVID-19 and puts them at risk of many complications such as pressure sores, contractures, and joint immobility. It is especially challenging to manage the positioning of patients with COVID-19, given that multiple devices and equipment surround the bed and that medical staff are at greater risk of infection with increased frequency of patient contact. A robotized hospital bed [47] was designed with a flexible mattress and an easily sanitized structure that allows the mobilization of major joints. This approach has several benefits. First, the effective mobilization of patients prevented the accumulation of secretions in the lungs. A self-movable bed that inclines can counter mucus engorgement and subsequent atelectasis. Second, the robotized system reduced the workload for health care workers (nurses or physiotherapists) and reduced the use and cost of disposable personal protective equipment. Third, passive mobilization of the major joints and muscles reduced pressure sores, venous thromboembolism, and muscle wasting.

Robots With Other Potential Roles in Critical Care

In patient evaluation, a KINARM robot (BKIN Technologies Ltd) [48-50] assessed the neurological outcomes of patients. In patients of postcardiac arrest, it accurately and precisely quantified neurological recovery, unlike conventional 5-point rating scales [48]. Similarly, it was able to better quantify neurocognitive impairment in terms of attention, executive function, and visuomotor function in patients of acute kidney injury [49] and patients of post-ICU discharge [50] as compared with Repeatable Battery for the Assessment of Neuropsychological Status, the standardized clinical assessment.

Mechanical compression [51] for cardiopulmonary resuscitation (CPR) using devices such as LUCAS II (Jolife AB) and Corpuls CPR (GS Elektromedizinische Geräte G Stemple GmbH) has been suggested. Although previous RCTs found that such mechanical devices have no clear advantage over manual CPR [52], we believe that mechanical compression technology can be included in future robots to enhance their capabilities. One such trial explored the use of robotic signal–guided CPR [53] to improve survival outcomes. Although no clear advantage of robots was found, this does not preclude further modifications and improvements to CPR-capable robots.

McSleepy [54] is an automated anesthesia drug delivery system for surgery. McSleepy administered appropriate drug doses by monitoring a patient's level of pain, muscle movements, and depth of consciousness. Although, as of yet, this has only been used in surgery, a closed-loop drug delivery system has tremendous potential for use in critical care settings, which requires constant and precise care.

An intelligent robotic hospital bed, Flexbed [55], with autonomous navigation ability, has been developed for the fast and safe transportation of patients of critical neurosurgery without needing to change beds. Preliminary trials in a simulated crowded hospital corridor environment showed its ability to transport patients quickly, safely, and efficiently while avoiding obstacles with a collision avoidance strategy.

Other robotic applications have been demonstrated in medical training, addressing a broad range of contexts and needs. In

XSL•FO
RenderX

pediatric care, a robotic simulator of premature neonates' wrists [56] was used to train novice caregivers to apply appropriate pressure, eliciting benefits in bone and muscle growth. The WKA-1R robot [57] is an airway simulator that accurately gauges the quality of intubation performance by providing a quantitative and objective determination.

Although this paper focused on patients who are critically ill, there are other potential ways in which robot use can be extended to critical care settings. In gait rehabilitation post spinal cord injury, a robot suit, Hybrid Assistive Limb (Cyberdyne Inc) [58] aided in recovering motor function and gait ability without increasing spasticity in individuals who are paraplegic and nonambulatory. Another robot, the Automatic Recovery Arm Motility Integrated System robot, is a dual exoskeleton robot designed specifically to help with paretic upper limb rehabilitation after stroke [59].

Future Research

Although many studies have uncovered the knowledge, attitudes, and perceptions of telepresence robots among health care staff, there were relatively few studies in this area that were conducted on patients themselves. Specifically, in the areas of patient confidentiality and privacy concerns, it might be pertinent to conduct more in-depth studies to uncover patient perspectives with regard to these issues.

In addition, all the included papers only compared the use of robots within the critical care setting before and after their implementation. To get a better idea of the extent to which robots specifically benefit critical care settings, more studies could be done directly comparing the use of robots within critical care versus noncritical care settings; for example, the cost or time benefit of a robot used within the ICU compared with the robot's use in a normal hospital ward.

As mentioned above, robots are moving toward being able to handle soft skills such as providing comfort. Other than Calmer, there are similar robots currently used outside the ICU, such as Paro [60]. Further development of such robots would be beneficial, especially given that patients in the ICU are more ill and isolated and might require more psychological support.

Currently, little regulation and few protocols exist for the use of robots, despite telepresence having existed for more than a decade. A paper by Clark et al [61] highlighted the types of cyberattacks on robots and the lack of current literature on the economic analysis of cyberattacks on robots. In another study, a protocol was created for the use of Pudu [30], which carefully considered the appropriate and practical use of Pudu in mental health care for isolated patients with COVID-19. It included aspects such as practical frameworks on patient interaction and robot movement, ethical and legal aspects of telecare, and cleaning and disinfection procedures. Although this protocol was newly drawn up for Pudu, it highlights and paves the way for similar protocols and frameworks for worldwide telepresence use. We hope that with more research in this area, suitable regulations and protocols can be implemented to address implementation issues such as manpower replacement in health care, cybersecurity issues, and subsequent ethical and legal consequences.

The papers we found did not mention robots as physician assistants. A physician assistant can accompany a physician to aid in decision-making, diagnosis and interpretation of signs, investigations, or management. Systems that incorporate both AI and robots within the critical setting could allow robots to act as physician assistants. For example, existing AI technology used for the early detection of sepsis [62] can potentially be incorporated into robots. Outside of medicine, many commercial companies are already moving toward incorporating AI into robots. Tesla Inc recently announced its intention to create a humanoid robot that could be used to replace *dangerous*, *repetitive*, *boring* tasks [63]. Although much progress is still to be made, it sets a bold tone for future robots that could also be extended to the medical field.

Limitations of This Review

Our review had several limitations. First, the aim of this project was to identify the different types of robots currently available for use in critical care settings. Currently, there are no theoretical frameworks to classify the types of robots; hence, we categorized the identified robots according to their functionality. However, as the types of robots varied widely, even within each theme, our review could only cover various types of robots in greater breadth rather than depth. In the future, with a larger volume of data, further research could perform detailed comparisons within each functional theme.

Second, we did not include conference abstracts and gray literature in our results as we felt that they did not have sufficient information for us to truly review their benefits and limitations. In addition, the papers were sourced from only 4 databases, which we decided to be the most relevant for the project.

Third, there were limitations in the studies themselves. For example, most studies lacked detailed economic analyses. Parameters such as cost savings, ICU occupancy, and staffing hours were dependent on the existing unique factors and circumstances within each ICU. There were also differences across studies in terms of what was included or omitted during these cost-benefit analyses. For instance, the components that went into calculating cost savings differed: some papers included robot maintenance fees in the overall value, but some did not. Therefore, we were unable to present a generalizable model that could be extrapolated to predict the amount of benefit for all ICUs.

Conclusions

Robotic use in critical care settings has been rising over the years. In particular, with the current COVID-19 pandemic, there has been greater emphasis on robot use in the ICU, as it allows efficient, safe, and quality contactless care. Although we initially set out believing that robots are more inclined toward aiding in mundane, repetitive tasks, we have discovered that they are capable of delivering substantial value in other more complicated aspects of patient care, including providing superior patient evaluation and rehabilitation. It was interesting to discover robotic use in addressing the softer aspects of patient care through examples such as the Pudu robot and Calmer, which have both proven to be well-integrated and positively perceived in their roles.

XSL•FO
RenderX

However, there are certain barriers that exist to robotic implementation in ICUs. We also hope that our paper will prompt the development of medicolegal frameworks for robotic use, especially in terms of sensitive aspects of care such as patient privacy or medical errors, and in other areas regarding the impact of robotic use, such as job employment.

Overall, given the present roles of robots and many other promising applications, we believe that there is a great opportunity for the further development of robotic technology for critical care, either alone or in combination with AI. If technical, financial, ethical, and legislative barriers to robotic use can be overcome, it would only be a matter of time before robotic presence in critical care becomes ubiquitous.

Acknowledgments

No financial support was obtained for the study.

Authors' Contributions

KCS, RT, and YD designed this study. RT and YD contributed to the data collection. RT and YD extracted data. RT and YD performed data analyses. RT and YD wrote the manuscript's initial draft. All 3 authors (RT, YD, and KCS) contributed to the revision of the manuscript. All authors have read and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 31 KB - jmir_v24i5e33380_app1.docx]

References

1. George E, Brand T, LaPorta A, Marescaux J, Satava R. Origins of robotic surgery: from skepticism to standard of care. J Soc Laproscopic Robot Surgeons 2018;22(4):e2018.00039. [doi: 10.4293/jsls.2018.00039]
2. Arabi Y, Murthy S, Webb S. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med 2020 May;46(5):833-836 [FREE Full text] [doi: 10.1007/s00134-020-05955-1] [Medline: 32125458]
3. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
4. da Costa Santos CM, de Mattos Pimenta CA, Nobre MR. The PICO strategy for the research question construction and evidence search. Rev Lat Am Enfermagem 2007 Jun;15(3):508-511 [FREE Full text] [doi: 10.1590/s0104-11692007000300023] [Medline: 17653438]
5. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Education and Research Archive. URL: https://era.library.ualberta.ca/items/48b9b989-c221-4df6-9e35-af782082280e [accessed 2022-04-11]
6. Adcock AK, Kosiorek H, Parikh P, Chauncey A, Wu Q, Demaerschalk BM. Reliability of robotic telemedicine for assessing critically ill patients with the full outline of UnResponsiveness score and Glasgow coma scale. Telemed J E Health 2017 Jul;23(7):555-560 [FREE Full text] [doi: 10.1089/tmj.2016.0225] [Medline: 28085631]
7. Alnobani O, Zakaria N, Temsah M, Jamal AA, Alkamel N, Tharkar S. Knowledge, attitude, and perception of health care personnel working in intensive care units of mass gatherings toward the application of telemedicine robotic remote-presence technology: a cross-sectional multicenter study. Telemed J E Health 2021 Dec 01;27(12):1423-1432. [doi: 10.1089/tmj.2020.0469] [Medline: 33691077]
8. Amodeo I, Pesenti N, Raffaeli G, Sorrentino G, Zorz A, Traina S, et al. Robotic therapy: cost, accuracy, and times. New challenges in the neonatal intensive care unit. Front Pharmacol 2019;10:1431 [FREE Full text] [doi: 10.3389/fphar.2019.01431] [Medline: 31849676]
9. Becevic M, Clarke M, Alnijoumi M, Sohal HS, Boren SA, Kim MS, et al. Robotic telepresence in a medical intensive care unit--clinicians' perceptions. Perspect Health Inf Manag 2015;12:1c [FREE Full text] [Medline: 26396554]
10. Bettinelli M, Lei Y, Beane M, Mackey C, Liesching TN. Does robotic telerounding enhance nurse-physician collaboration satisfaction about care decisions? Telemed J E Health 2015 Aug;21(8):637-643. [doi: 10.1089/tmj.2014.0162] [Medline: 25938296]
11. Burke RV, Berg BM, Vee P, Morton I, Nager A, Neches R, et al. Using robotic telecommunications to triage pediatric disaster victims. J Pediatr Surg 2012 Jan;47(1):221-224. [doi: 10.1016/j.jpedsurg.2011.10.046] [Medline: 22244422]
12. Duan S, Liu L, Chen Y, Yang L, Zhang Y, Wang S, et al. A 5G-powered robot-assisted teleultrasound diagnostic system in an intensive care unit. Crit Care 2021 Apr 07;25(1):134 [FREE Full text] [doi: 10.1186/s13054-021-03563-z] [Medline: 33827638]

XSL•FO
RenderX

13. Frazzitta G, Zivi I, Valsecchi R, Bonini S, Maffia S, Molatore K, et al. Effectiveness of a very early stepping verticalization protocol in severe acquired brain injured patients: a randomized pilot study in ICU. PLoS One 2016;11(7):e0158030 [FREE Full text] [doi: 10.1371/journal.pone.0158030] [Medline: 27447483]
14. Garingo A, Friedlich P, Tesoriero L, Patil S, Jackson P, Seri I. The use of mobile robotic telemedicine technology in the neonatal intensive care unit. J Perinatol 2012 Jan 26;32(1):55-63. [doi: 10.1038/jp.2011.72] [Medline: 21617643]
15. Garingo A, Friedlich P, Chavez T, Tesoriero L, Patil S, Jackson P, et al. "Tele-rounding" with a remotely controlled mobile robot in the neonatal intensive care unit. J Telemed Telecare 2015 Jun 26;22(2):132-138. [doi: 10.1177/1357633x15589478]
16. Goldberg MH, Bily T, Ritter G, Barrera R. Economic benefit of a novel intensivist coverage in a surgical ICU. ICU Director 2013 Oct 18;4(6):291-295. [doi: 10.1177/1944451613507450]
17. Holsti L, MacLean K, Oberlander T, Synnes A, Brant R. Calmer: a robot for managing acute pain effectively in preterm infants in the neonatal intensive care unit. Pain Report 2019 Mar;4(2):e727. [doi: 10.1097/pr9.0000000000000727]
18. Holt T, Sari N, Hansen G, Bradshaw M, Prodanuk M, McKinney V, et al. Remote presence robotic technology reduces need for pediatric interfacility transportation from an isolated northern community. Telemed J E Health 2018 Nov;24(11):927-933. [doi: 10.1089/tmj.2017.0211] [Medline: 29394155]
19. Ito K, Tsuruta K, Sugano S, Iwata H. Evaluation of a wearable tele-echography robot system: FASTele in a vehicle using a mobile network. In: Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 30 - Sep 3, 2011; Boston, MA, USA. [doi: 10.1109/iembs.2011.6090389]
20. Lazzara E, Benishek L, Patzer B, Gregory ME, Hughes AM, Heyne K, et al. Utilizing telemedicine in the trauma intensive care unit: does it impact teamwork? Telemed J E Health 2015 Aug;21(8):670-676. [doi: 10.1089/tmj.2014.0074] [Medline: 25885369]
21. Marini CP, Ritter G, Sharma C, McNelis J, Goldberg M, Barrera R. The effect of robotic telerounding in the surgical intensive care units impact on medical education. J Robot Surg 2015 Mar 8;9(1):51-56. [doi: 10.1007/s11701-014-0489-5] [Medline: 26530971]
22. Marttos A, Kelly E, Graygo J, Rothenberg P, Alonso G, Kuchkarian FM, et al. Usability of telepresence in a level 1 trauma center. Telemed J E Health 2013 Apr;19(4):248-251. [doi: 10.1089/tmj.2012.0102] [Medline: 23509919]
23. McNelis J, Schwall G, Collins J. Robotic remote presence technology in the surgical intensive care unit. J Trauma Acute Care Surg 2012 Feb;72(2):527-530. [doi: 10.1097/TA.0b013e31822f7d3b] [Medline: 22327990]
24. Murray C, Ortiz E, Kubin C. Application of a robot for critical care rounding in small rural hospitals. Crit Care Nurs Clin North Am 2014 Dec;26(4):477-485. [doi: 10.1016/j.ccell.2014.08.006] [Medline: 25438890]
25. Prokazova PR, Piradov MA, Ryabinkina YV, Kunzevich GI, Gnedovskaya EV, Popova LA. Robot-assisted therapy using the MOTOmed letto 2 for the integrated early rehabilitation of stroke patients admitted to the intensive care unit. Hum Physiol 2017 Feb 16;42(8):885-890. [doi: 10.1134/s0362119716080107]
26. Reynolds EM, Grujovski A, Wright T, Foster M, Reynolds HN. Utilization of robotic "remote presence" technology within North American intensive care units. Telemed J E Health 2012 Sep;18(7):507-515. [doi: 10.1089/tmj.2011.0206] [Medline: 22738430]
27. Rincon F, Vibbert M, Childs V, Fry R, Caliguri D, Urtecho J, et al. Implementation of a model of robotic tele-presence (RTP) in the neuro-ICU: effect on critical care nursing team satisfaction. Neurocrit Care 2012 Aug 1;17(1):97-101. [doi: 10.1007/s12028-012-9712-2] [Medline: 22547040]
28. Rocca A, Pignat J, Berney L, Jöhr J, Van de Ville D, Daniel RT, et al. Sympathetic activity and early mobilization in patients in intensive and intermediate care with severe brain injuries: a preliminary prospective randomized study. BMC Neurol 2016 Sep 13;16:169 [FREE Full text] [doi: 10.1186/s12883-016-0684-2] [Medline: 27619015]
29. Rogove HJ, McArthur D, Demaerschalk BM, Vespa PM. Barriers to telemedicine: survey of current users in acute care units. Telemed J E Health 2012 Jan;18(1):48-53. [doi: 10.1089/tmj.2011.0071] [Medline: 22082107]
30. Ruiz-del-Solar J, Salazar M, Vargas-Araya V, Campodonico U, Marticorena N, Pais G, et al. Mental and emotional health care for COVID-19 patients: employing pudu, a telepresence robot. IEEE Robot Automat Mag 2021 Mar;28(1):82-89. [doi: 10.1109/mra.2020.3044906]
31. Shimizu N, Motomura M, Saito O, Ikeyama T. Affinity of health care providers for artificially intelligent robots at bedside in the pediatric intensive care unit. Chiba Med J 2019;95(E)(2):17-19. [doi: 10.20776/S03035476-95E-2-P17]
32. Sucher J, Todd S, Jones S, Throckmorton T, Turner K, Moore F. Robotic telepresence: a helpful adjunct that is viewed favorably by critically ill surgical patients. Am J Surg 2011 Dec;202(6):843-847. [doi: 10.1016/j.amjsurg.2011.08.001] [Medline: 22137142]
33. Summerfield M, Seagull F, Vaidya N, Xiao Y. Use of pharmacy delivery robots in intensive care units. Am J Health Syst Pharm 2011 Jan 01;68(1):77-83. [doi: 10.2146/ajhp100012] [Medline: 21164071]
34. Vespa PM, Miller C, Hu X, Nenov V, Buxey F, Martin NA. Intensive care unit robotic telepresence facilitates rapid physician response to unstable patients and decreased cost in neurointensive care. Surg Neurol 2007 Apr;67(4):331-337. [doi: 10.1016/j.surneu.2006.12.042] [Medline: 17350395]
35. Wang J, Peng C, Zhao Y, Ye R, Hong J, Huang H, et al. Application of a robotic tele-echography system for COVID-19 pneumonia. J Ultrasound Med 2021 Feb 29;40(2):385-390. [doi: 10.1002/jum.15406] [Medline: 32725833]

XSL•FO
RenderX

36. Williams N, MacLean K, Guan L, Collet JP, Holsti L. Pilot testing a robot for reducing pain in hospitalized preterm infants. OTJR (Thorofare N J) 2019 Apr 15;39(2):108-115. [doi: 10.1177/1539449218825436] [Medline: 30770034]
37. Ye R, Zhou X, Shao F, Xiong L, Hong J, Huang H, et al. Feasibility of a 5G-based robot-assisted remote ultrasound system for cardiopulmonary assessment of patients with Coronavirus disease 2019. Chest 2021 Jan;159(1):270-281 [FREE Full text] [doi: 10.1016/j.chest.2020.06.068] [Medline: 32653568]
38. Zeiler F, Smielewski P. Application of robotic transcranial Doppler for extended duration recording in moderate/severe traumatic brain injury: first experiences. Crit Ultrasound J 2018 Jul 23;10(1):16 [FREE Full text] [doi: 10.1186/s13089-018-0097-0] [Medline: 30033501]
39. Hu M, Hsu S, Yip P, Jeng J, Wang Y. Early and intensive rehabilitation predicts good functional outcomes in patients admitted to the stroke intensive care unit. Disabil Rehabil 2010 Feb 04;32(15):1251-1259. [doi: 10.3109/09638280903464448] [Medline: 20131942]
40. Tan BY, Kanneganti A, Lim LJ, Tan M, Chua YX, Tan L, et al. Burnout and associated factors among health care workers in Singapore during the COVID-19 pandemic. J Am Med Dir Assoc 2020 Dec;21(12):1751-8.e5 [FREE Full text] [doi: 10.1016/j.jamda.2020.09.035] [Medline: 33256955]
41. Choi HK, Cui C, Seok H, Bae J, Jeon JH, Lee GE, et al. Feasibility of ultraviolet light-emitting diode irradiation robot for terminal decontamination of coronavirus disease 2019 (COVID-19) patient rooms. Infect Control Hosp Epidemiol 2022 Feb 09;43(2):232-237 [FREE Full text] [doi: 10.1017/ice.2021.95] [Medline: 33685546]
42. Lorca-Oró C, Vila J, Pleguezuelos P, Vergara-Alert J, Rodon J, Majó N, et al. Rapid SARS-CoV-2 inactivation in a simulated hospital room using a mobile and autonomous robot emitting ultraviolet-C light. J Infect Dis 2022 Feb 15;225(4):587-592 [FREE Full text] [doi: 10.1093/infdis/jiab551] [Medline: 34904659]
43. Freeman WD, Sanghavi DK, Sarab MS, Kindred MS, Dieck EM, Brown SM, et al. Robotics in simulated COVID-19 patient room for health care worker effector tasks: preliminary, feasibility experiments. Mayo Clin Proc Innov Qual Outcomes 2021 Feb;5(1):161-170 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2020.12.005] [Medline: 33521585]
44. Vagvolgyi BP, Khrenov M, Cope J, Deguet A, Kazanzides P, Manzoor S, et al. Telerobotic operation of intensive care unit ventilators. Front Robot AI 2021 Jun 24;8:612964 [FREE Full text] [doi: 10.3389/frobt.2021.612964] [Medline: 34250025]
45. Wazir HK, Lourido C, Chacko SM, Kapila V. A COVID-19 emergency response for remote control of a dialysis machine with mobile HRI. Front Robot AI 2021 May 7;8:612855 [FREE Full text] [doi: 10.3389/frobt.2021.612855] [Medline: 34262944]
46. Ort T, Hanumara N, Antonini A, Araki B, Abu-Khalaf M, Detienne M, et al. MIT emergency-vent: an automated resuscitator bag for the COVID-19 crisis. Annu Int Conf IEEE Eng Med Biol Soc 2021 Nov;2021:4998-5004. [doi: 10.1109/EMBC46164.2021.9630882] [Medline: 34892330]
47. Ceccarelli M, Sarmati L, Ambrogi V. A robotized hospital bed for COVID-19 patients in intensive care treatments. In: Proceedings of the 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA). 2021 Presented at: 2021 IEEE International Conference on Automation/XXIV Congress of the Chilean Association of Automatic Control (ICA-ACCA); Mar 22-26, 2021; Valparaíso, Chile. [doi: 10.1109/icaacca51523.2021.9465310]
48. Boyd JG, Scott S. 538 Robotic technology identifies neurocognitive deficits in high functioning cardiac arrest survivors. Crit Care Med 2015;43:136. [doi: 10.1097/01.ccm.0000474366.26395.31]
49. Vanderlinden J, Semrau J, Silver S, Holden R, Scott S, Boyd J. Acute kidney injury is associated with subtle but quantifiable neurocognitive impairments. Nephrol Dial Transplant 2022 Jan 25;37(2):285-297. [doi: 10.1093/ndt/gfab161] [Medline: 33881540]
50. Wood MD, Maslove DM, Muscedere J, Scott SH, Boyd JG, Canadian Critical Care Trials Group. Robotic technology provides objective and quantifiable metrics of neurocognitive functioning in survivors of critical illness: a feasibility study. J Crit Care 2018 Dec;48:228-236. [doi: 10.1016/j.jcrc.2018.09.011] [Medline: 30243203]
51. Eichhorn S, Mendoza A, Prinzing A, Stroh A, Xinghai L, Polski M, et al. Corpuls CPR generates higher mean arterial pressure than LUCAS II in a pig model of cardiac arrest. Biomed Res Int 2017;2017:5470406-5470409 [FREE Full text] [doi: 10.1155/2017/5470406] [Medline: 29392137]
52. Nolan JP, Soar J, Zideman DA, Biarent D, Bossaert LL, Deakin C, ERC Guidelines Writing Group. European resuscitation council guidelines for resuscitation 2010 section 1. Executive summary. Resuscitation 2010 Oct;81(10):1219-1276. [doi: 10.1016/j.resuscitation.2010.08.021] [Medline: 20956052]
53. Salcido D, Sundermann M, Koller A, Flickinger K, Menegazzi J. Abstract 19337: preliminary evaluation of a pressor-enhanced approach to robotic signal-guided cardiopulmonary resuscitation in a swine model of ventricular fibrillation. Circulation 2016;134(1).
54. Wehbe M, Arbeid E, Cyr S, Mathieu PA, Taddei R, Morse J, et al. A technical description of a novel pharmacological anesthesia robot. J Clin Monit Comput 2014 Feb 15;28(1):27-34. [doi: 10.1007/s10877-013-9451-8] [Medline: 23771652]
55. Wang C, Savkin AV, Clout R, Nguyen HT. An intelligent robotic hospital bed for safe transportation of critical neurosurgery patients along crowded hospital corridors. IEEE Trans Neural Syst Rehabil Eng 2015 Sep;23(5):744-754. [doi: 10.1109/tnsre.2014.2347377]

XSL•FO
RenderX

56. Adnan K, Ahmad I, Coussens M, Eliakim A, Gallitto S, Grochow D, et al. A haptic simulator for training the application of range of motion exercise to premature infants. J Med Device 2009;3(4):041008. [doi: 10.1115/1.4000430]
57. Noh Y, Segawa M, Shimomura A, Ishii H, Solis J, Hatake K, et al. WKA-1R Robot assisted quantitative assessment of airway management. Int J CARS 2008 Sep 3;3(6):543-550. [doi: 10.1007/s11548-008-0238-1]
58. Watanabe H, Marushima A, Kawamoto H, Kadone H, Ueno T, Shimizu Y, et al. Intensive gait treatment using a robot suit hybrid assistive limb in acute spinal cord infarction: report of two cases. J Spinal Cord Med 2019 May;42(3):395-401 [FREE Full text] [doi: 10.1080/10790268.2017.1372059] [Medline: 28990874]
59. Pignolo L, Dolce G, Basta G, Lucca L, Serra S, Sannita W. Upper limb rehabilitation after stroke: ARAMIS a "robo-mechatronic" innovative approach and prototype. In: Proceedings of the 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob). 2012 Presented at: 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob); Jun 24-27, 2012; Rome, Italy. [doi: 10.1109/biorob.2012.6290868]
60. Hung L, Liu C, Woldum E, Au-Yeung A, Berndt A, Wallsworth C, et al. The benefits of and barriers to using a social robot PARO in care settings: a scoping review. BMC Geriatr 2019 Aug 23;19(1):232 [FREE Full text] [doi: 10.1186/s12877-019-1244-6] [Medline: 31443636]
61. Clark G, Doran M, Andel T. Cybersecurity issues in robotics. In: Proceedings of the 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA). 2017 Presented at: 2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA); Mar 27-31, 2017; Savannah, GA, USA. [doi: 10.1109/cogsima.2017.7929597]
62. Kalil A, Dias VD, Rocha CD, Morales H, Fressatto J, Faria R. Sepsis risk assessment: a retrospective analysis after a cognitive risk management robot (Robot Laura®) implementation in a clinical-surgical unit. Res Biomed Eng 2018 Nov 22;34(4):310-316. [doi: 10.1590/2446-4740.180021]
63. Ryan J. Elon Musk and Tesla: An Electrifying Love Affair. Thousand Oaks, California: SAGE Publications; 2021.

Abbreviations

AI: artificial intelligence
CPR: cardiopulmonary resuscitation
FAST: focused assessment with sonography for trauma
ICU: intensive care unit
NICU: neonatal intensive care unit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial
TCD: transcranial Doppler

Edited by G Eysenbach; submitted 05.09.21; peer-reviewed by C Sutherland, K Yap; comments to author 02.12.21; revised version received 22.01.22; accepted 06.03.22; published 16.05.22.

Please cite as:
Teng R, Ding Y, See KC
Use of Robots in Critical Care: Systematic Review
J Med Internet Res 2022;24(5):e33380
URL: https://www.jmir.org/2022/5/e33380
doi:10.2196/33380
PMID:35576567

©Rachel Teng, Yichen Ding, Kay Choong See. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.05.2022. 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.

XSL•FO
RenderX

Review

Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review

Robert Jakob[1], MSc; Samira Harperink[2], BA; Aaron Maria Rudolf[2], MD; Elgar Fleisch[1,2,3], PhD; Severin Haug[4], PhD; Jacqueline Louise Mair[3,5], PhD; Alicia Salamanca-Sanabria[3], PhD; Tobias Kowatsch[1,2,3], PhD

[1]Centre for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

[2]Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

[3]Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence And Technological Enterprise, Singapore, Singapore

[4]Swiss Research Institute for Public Health and Addiction, Zurich University, Zurich, Switzerland

[5]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

Corresponding Author:
Robert Jakob, MSc
Centre for Digital Health Interventions
Department of Management, Technology and Economics
ETH Zurich
Professur Informationsmanagement, WEV G 217
Weinbergstr. 56/58
Zurich, 8092
Switzerland
Phone: 41 44 632 53 57
Email: rjakob@ethz.ch

Abstract

Background: Mobile health (mHealth) apps show vast potential in supporting patients and health care systems with the increasing prevalence and economic costs of noncommunicable diseases (NCDs) worldwide. However, despite the availability of evidence-based mHealth apps, a substantial proportion of users do not adhere to them as intended and may consequently not receive treatment. Therefore, understanding the factors that act as barriers to or facilitators of adherence is a fundamental concern in preventing intervention dropouts and increasing the effectiveness of digital health interventions.

Objective: This review aimed to help stakeholders develop more effective digital health interventions by identifying factors influencing the continued use of mHealth apps targeting NCDs. We further derived quantified adherence scores for various health domains to validate the qualitative findings and explore adherence benchmarks.

Methods: A comprehensive systematic literature search (January 2007 to December 2020) was conducted on MEDLINE, Embase, Web of Science, Scopus, and ACM Digital Library. Data on intended use, actual use, and factors influencing adherence were extracted. Intervention-related and patient-related factors with a positive or negative influence on adherence are presented separately for the health domains of NCD self-management, mental health, substance use, nutrition, physical activity, weight loss, multicomponent lifestyle interventions, mindfulness, and other NCDs. Quantified adherence measures, calculated as the ratio between the estimated intended use and actual use, were derived for each study and compared with the qualitative findings.

Results: The literature search yielded 2862 potentially relevant articles, of which 99 (3.46%) were included as part of the inclusion criteria. A total of 4 intervention-related factors indicated positive effects on adherence across all health domains: personalization or tailoring of the content of mHealth apps to the individual needs of the user, reminders in the form of individualized push notifications, user-friendly and technically stable app design, and personal support complementary to the digital intervention. Social and gamification features were also identified as drivers of app adherence across several health domains. A wide variety of patient-related factors such as user characteristics or recruitment channels further affects adherence. The derived adherence scores of the included mHealth apps averaged 56.0% (SD 24.4%).

Conclusions: This study contributes to the scarce scientific evidence on factors that positively or negatively influence adherence to mHealth apps and is the first to quantitatively compare adherence relative to the intended use of various health domains. As

XSL•FO
RenderX

underlying studies mostly have a pilot character with short study durations, research on factors influencing adherence to mHealth apps is still limited. To facilitate future research on mHealth app adherence, researchers should clearly outline and justify the app's intended use; report objective data on actual use relative to the intended use; and, ideally, provide long-term use and retention data.

(J Med Internet Res 2022;24(5):e35371) doi:10.2196/35371

KEYWORDS

intended use; adherence; engagement; attrition; retention; mHealth; eHealth; digital health intervention; noncommunicable disease; NCD; mobile phone

Introduction

Rationale

Digital health interventions (DHIs) show vast potential in supporting patients and health care systems with the globally increasing prevalence and economic costs of noncommunicable diseases (NCDs), which is the leading causes of death and disability worldwide [1,2]. More specifically, mobile health (mHealth) apps are now considered accessible and scalable solutions to promoting behavior change among patients, improving health outcomes, and reducing health care costs [3-5]. Correspondingly, the number of available mHealth apps has continuously grown to >300.000, with approximately 200 new mHealth apps released each day [2,6].

However, despite increasing evidence and availability, mHealth apps are subject to significant dropout rates, with a substantial proportion of users not adhering to them as intended [7,8]. Recent research has shown that up to 80% of all participants in mHealth interventions only engage at a minimum level, do not log into the mHealth app more than once, and do not consistently use the app in the long term [9]. Another study examining mHealth app use in more extensive real-world settings reported low retention rates, with only 3.9% of participants using mHealth apps for >15 days [10]. The reported low adherence and high attrition levels further highlight the necessity of developing more effective models, best practices, and interventions [8,11].

As nonadherence relative to intended use jeopardizes treatment success and, thus, might lead to an increased number of hospitalizations, it is considered a fundamental concern in the development of mHealth apps [8,12-15]. However, the scientific body of literature lacks concise conceptualizations and measures for the intended use of mHealth apps, whereas intervention components and factors influencing adherence remain to be explored [13,16]. Following previous studies, we define *adherence* as "the degree to which the user followed the program as it was designed," which can be paraphrased as "adherence relative to the intended use" [13,17,18].

With smartphone apps being the primary intervention component, adherence relative to the intended use is principally informed by user acceptance and the use of information technology [19]. Previous research underscores the necessity of mHealth apps that must be first accepted and used in an intended way to then achieve a desired health behavior change [3,19]. Correspondingly, previous research has identified factors affecting the uptake of and engagement with health and well-being smartphone apps [11,20-23]. Many of these strategies and factors, such as well-designed reminders, self-monitoring features, and embedded health professional support, have been applied across various health domains [11,20,21]. Some of these factors, such as reminders, can be further applied as retention methods and strategies for cohort studies in general and may thus extend the scope of DHIs [24,25].

Identifying the factors that influence adherence relative to intended use may support and extend these findings. Given previous research and their relation to technology use and acceptance, we can assume that these factors may be not only generalizable across various health behavior domains but also be applicable to DHIs using alternative information technologies. To our knowledge, no systematic review has been conducted on the factors influencing adherence to mHealth apps designed to prevent or manage NCDs. Furthermore, to the best of our knowledge, no review has previously explored the quantifying of adherence to assess qualitatively identified factors.

Objectives

Preventing intervention dropouts and thus increasing the effectiveness of mHealth apps requires an understanding of the factors that act as barriers to or facilitators of intervention adherence. This review aimed to identify factors influencing adherence relative to the intended use of mHealth apps, which may help stakeholders better plan, develop, and evaluate mHealth apps. To help readers navigate through the identified factors, we further categorized them into intervention-related factors that app developers can potentially improve upon through product changes (eg, the inclusion of certain app features) and patient-related factors that are hardly adjustable (eg, user characteristics). These factors were separated into their potential positive or negative influences on adherence.

In the absence of a universally agreed-upon approach to measuring adherence to mHealth interventions, we exploratively derived an adherence score as the ratio between the intended and actual use of each study to describe adherence quantitatively and consistently. The primary aim of the resulting adherence score was to quantitatively assess the findings from the qualitative extraction of factors influencing adherence. As the intended use varies substantially across different mHealth apps, we extracted the intended use for each included mHealth app individually. We then compared the intended use with the actual use reported in the corresponding study. To the best of our knowledge, this exploratory approach of a quantified adherence score has not been applied previously.

XSL•FO
RenderX

In summary, this review aimed to answer the following research questions:

1. Which intervention-related factors influence adherence relative to the intended use of mHealth apps targeting NCDs in adults?
2. Which patient-related factors influence adherence relative to the intended use of mHealth apps targeting NCDs in adults?
3. How do the adherence rates of mHealth apps for NCDs compare across different health domains?

Methods

Database Selection and Search Strategy

This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1). A review protocol was submitted to the Federal Office of Public Health of the Swiss Confederation on October 7, 2020, but was not publicly registered.

The electronic databases Embase (including MEDLINE and PubMed), Web of Science, Scopus, and ACM Digital Library were searched using a predefined search strategy that included search terms related to mHealth apps, app use, and study design (Multimedia Appendix 2). The search terms were customized for each electronic database, and if the respective database allowed it, the corresponding Medical Subject Heading terms or topics were also integrated. Articles published in English between June 2007 (release of the iPhone) and December 2020, which focused on adult populations, were included. Studies that focused on communicable diseases were excluded. The inclusion and exclusion criteria listed in Textbox 1 were used to identify relevant articles.

Textbox 1. List of eligibility criteria (Population, Intervention, Comparison, Outcomes, and Study component [PICOS] along with inclusion and exclusion criteria and applied filters).

Inclusion criteria

- Participants: adults aged ≥18 years; studies that included individuals aged ≥16 years were included if at least 70% of the participants were aged ≥18 years
- Intervention and context: studies investigating digital interventions that aimed to change ≥1 health behavior and the stated goal of the intervention was to prevent or treat a noncommunicable disease or condition
- Comparison: any kind of comparison
- Outcomes
  - Qualitative: factors predicting adherence or nonadherence relative to the intended use
  - Quantitative: information on the actual and intended use of the intervention or information on adherence relative to the intended use
- Study design: primary and secondary studies, including randomized controlled trials, systematic reviews, meta-analyses, observational studies, single-center experiments, feasibility studies, pilot studies, and experimental studies

Exclusion criteria

- Participants: children and adolescents aged <18 years and animals
- Intervention and context:
  - Studies with the smartphone not being the primary intervention component
  - Interventions not targeting noncommunicable diseases; for example, communicable diseases (influenza, norovirus, Ebola, and COVID-19)
- Comparison: none
- Outcomes: the study does not contain information on the actual and intended use of the intervention
- Study design: animal and laboratory studies, case reports, case series, narrative reviews, expert opinions, editorials, conference abstracts, and study protocols

Applied filters

- Time: studies published from June 2007 onward
- Language: English
- Access: open access or via institutional log-in

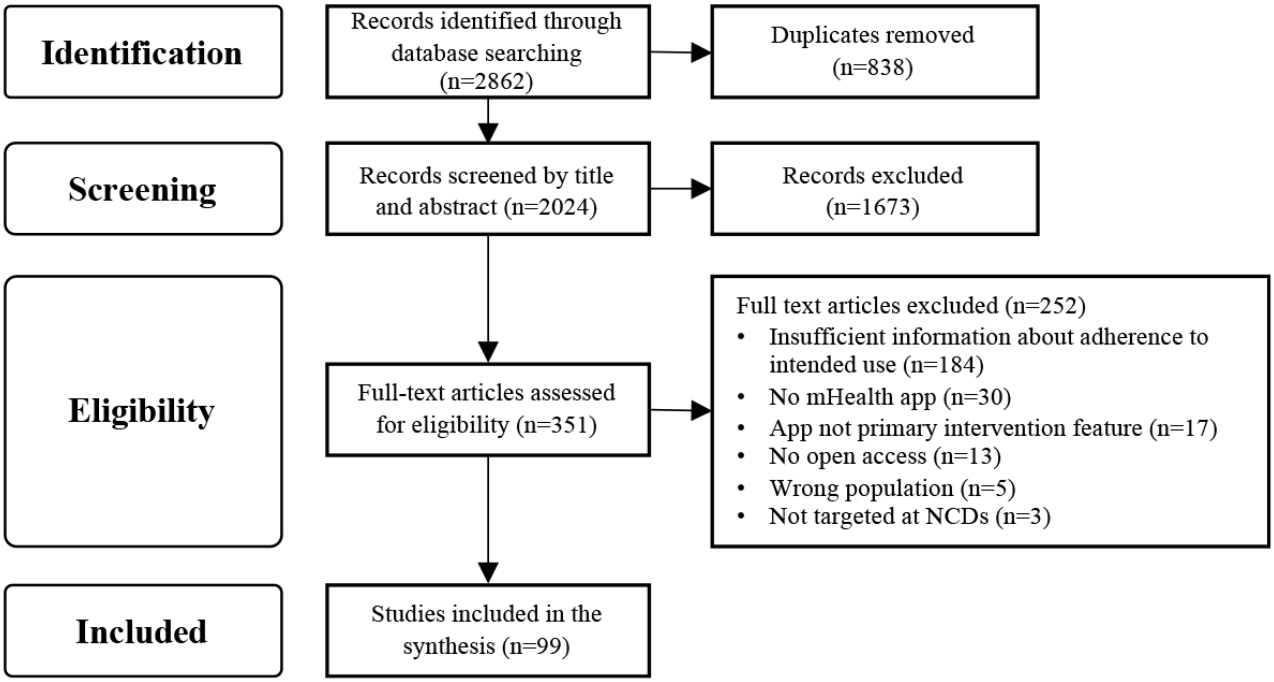
Screening Process and Eligibility Criteria

The selection of publications was conducted in several steps (Figure 1). First, potentially relevant publications were identified by searching the literature databases. After excluding duplicates, titles and abstracts were independently reviewed by 3 researchers (SH, RJ, and AMR) according to the inclusion and exclusion criteria listed in Textbox 1. Disagreements were resolved through discussion. In a second screening step, the full texts of relevant articles were independently reviewed by 4 researchers (SH, RJ, AMR, and JLM) concerning the fulfillment of the inclusion and exclusion criteria. Discrepancies were again

XSL•FO
RenderX

resolved through discussions. The web-based program Covidence (Veritas Health Innovation) was used for the entire process. After applying the search strategies outlined in Multimedia Appendix 2, the resulting database reference lists were imported into the Covidence database. The following Covidence features were used in the process: duplicate removal, title and abstract screening, full-text review, and export of PRISMA flowchart.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart illustrating the inclusion and exclusion of studies. mHealth: mobile health; NCD: noncommunicable disease.



Data Extraction

The following information was extracted for each included study: general study characteristics, study population characteristics, intervention characteristics, factors influencing adherence relative to intended use, and information on app use (Multimedia Appendix 3 [20,26-123]).

General study characteristics included the title, first author, year of publication, journal name, country, study design, and health domain.

Study population characteristics comprised age, gender, type of population (clinical or general population), type of disease, and number of study participants.

Intervention characteristics were app name, smartphone operating system (universal, Android only, or iOS only), type of mHealth app offered (publicly available or research apps), app developer (private company or nonprofit organization), level of personal support (no personal support during the intervention or continuous personal support), external monetary incentives and their value in US dollars, intervention duration in days, and effectiveness of the intervention in terms of health outcomes.

Factors influencing adherence relative to intended use were extracted and characterized as intervention-related factors (factors that developers can potentially improve through product changes) or patient-related factors (factors that can hardly be influenced by app developers, such as user characteristics). These factors were further categorized based on their positive or negative influence on adherence.

Information on app use comprised intended app use, actual app use, the number of intended intervention interactions, and interaction frequency (eg, daily or weekly intended use). The adherence score was defined as *adherence relative to the intended use* and was derived as the quantified ratio of intended use to actual use.

Synthesis and Statistical Analyses

As a first step, the identified studies were categorized based on the mHealth app they investigated as follows: apps targeting NCD self-management (including the four main NCDs: asthma, diabetes, cardiovascular disease [CVD], and cancer), mental health disorders (anxiety and depression), substance use disorders (alcohol and tobacco), and behavioral risk factors (nutrition, physical activity, and weight loss). The categories were then further refined into the following health domains: NCD self-management (asthma management, diabetes management, CVD management, cancer management, and medication adherence), mental health (anxiety, depression, and multidisciplinary and others), substance use (alcohol, tobacco, and multidisciplinary and others), nutrition, physical activity, weight loss, multicomponent lifestyle interventions, and other NCDs.

In the second step, intervention-related and patient-related factors that were outlined in the studies as barriers to or facilitators of adherence were qualitatively evaluated,

XSL•FO
RenderX

summarized within the health domain, and categorized based on their positive or negative influence.

In the third step, the adherence score was derived for each study focusing directly on a specific mHealth app (97/99, 98%) and was calculated as the ratio of intended use to actual use. The mean adherence scores were calculated for each health domain.

In the fourth step, correlations of adherence scores with other extracted variables were examined, and where possible, the qualitative results from step 2 were quantitatively compared for each health domain. Quantitative analyses were performed using SPSS Statistics (version 27; IBM Corp). Correlations of adherence scores with continuous variables (eg, the average age of study participants) were calculated using the Pearson correlation. Correlations with ordinal variables (eg, level of personal support) were calculated using the Spearman correlation.

Finally, a list of universally relevant factors with recommendations for the development and evaluation of mHealth apps was developed.

Results

Selection and Inclusion of Studies

The search of electronic databases was performed on January 3, 2021, and yielded 2862 articles. After excluding duplicates, 70.72% (2024/2862) of publications remained for the title and abstract screening. Subsequently, the full texts of 17.34% (351/2024) of articles were examined. Of the 351 studies, 99 (28.2%) were finally included in the data synthesis. Figure 1 visualizes the selection process and reasons for exclusion.

Characteristics of Included Studies

In total, 99 studies were included in this review. Of these 99 studies, 2 (2%) were systematic reviews, and 97 (98%) evaluated specific mHealth apps. Randomized controlled trials were the most frequent study type (35/97, 36%), followed by pilot trials (31/97, 32%), pilot randomized controlled trials (12/97, 12%), cohort studies, mixed methods studies, and observational studies (each 5/97, 5%). Most of the studies were conducted in North America (37/97, 38%), followed by Europe (34/97, 35%), Australia (15/97, 15%), and Asia (10/97, 10%). Of the 97 studies, 90 (93%) were published within the past 5 years: 46 (47%) in 2020, 12 (12%) in 2019, 12 (12%) in 2018, 9 (9%) in 2017, 8 (8%) in 2016, 2 (2%) in 2015, 4 (4%) in 2014, and 4 (4%) in 2013. The mean intervention duration was 111.4 (SD 132; range 7-730) days, with 27% (26/97) of studies lasting 1 to 4 weeks, 46% (46/97) of studies lasting between 1 and 3 months, 18% (17/97) of studies lasting between 3 and 12 months, and 9% (7/97) of studies lasting longer than a year. In 32% (31/97) of studies, monetary incentives were provided to the participants as compensation. The mean derived incentive value was US $105.42 (SD US $18.65; range US $7 to US $430).

Characteristics of Study Populations

The total number of participants in the included studies evaluating specific mHealth apps was 72,046. The mean number of study participants was 750.5 (SD 2800.7; range 9-19,233). Of the 96 studies reporting exact participants numbers, 65 (68%) had <100 participants, 21 (22%) had 100 to 1000 participants, and only 10 (10%) studies had >1000 study participants. In several studies, most study participants had a pre-existing condition (82/97, 85%), with mental health conditions being the most prevalent (21/97, 22%), followed by obesity and being overweight (15/97, 15%), substance abuse (9/97, 9%), cancer (7/97, 7%), diabetes (5/97, 5%), CVD (5/97, 5%), and sleep disorder (4/97, 4%). In 15% (15/97) of studies, most participants were healthy. The overall mean age was 44.6 (SD 12.9; range 19.9-86) years, and the mean percentage of women was 62% (SD 22.8%; range 0%-100%).

Characteristics of mHealth Apps

Of the 97 reviewed apps, 50 (51%) were available for both iOS and Android. The remaining apps were exclusively available on either iOS or Android platforms (both 17/97, 18%). The authors of 13% (13/98) of studies did not clearly outline on which platforms the apps were distributed. Of the 97 reviewed apps, 47 (48%) were publicly available, whereas 52 (54%) were exclusively available to study participants. Approximately 38% (37/97) of apps were developed by private commercial companies (eg, software companies), and 64% (62/97) of apps were developed by nonprofit organizations (eg, academic institutes). Of the studies that clearly outlined their study procedure, 34% (32/93) included personal contact with health personnel during the study as an intervention component. In comparison, 66% (61/93) of the apps provided only personal support in the app onboarding phase. Of the 54 studies evaluating the app's effectiveness on a primary outcome, the authors of 34 (63%) studies highlighted their app as effective, and the authors of 20 (37%) studies highlighted their app as ineffective. The most common explanation of intended use, according to the authors, or derived from information on intervention design, was daily tracking (eg, daily diary entries; 36/97, 36%), followed by activity completion (eg, completion of a certain amount of coaching modules; 19/97, 20%), daily use (eg, daily log-in; 17/97, 18%), daily activity completion (6/97, 6%), weekly tracking (5/97, 5%), weekly use (4/97, 4%), activity completion+daily tracking (3/97, 3%), weekly use time (eg, using the app 1 hour per day; 2/97, 2%), prolonged use (eg, no inactivity for >2 weeks; 2/97, 2%), and biweekly tracking (1/97, 1%). Approximately 95% (92/96) of studies reported data on actual use based on objective app use data, and 4% (4/96) of studies reported data based on qualitative feedback from users. The mean adherence score across all interventions was 56.0% (SD 24.4%; range 2.6%-96.0%). The mean number of interactions within the study period amounted to 90.1 (SD 145.9; range 1-730) interactions. Of the 97 apps, 14 (14%) apps were intended for ≥2 daily interactions with the app, 53 (55%) apps were intended for 1 daily interaction, and 17 (18%) were intended for weekly interactions; in 13 (13%) apps, users only had to use the app once a month to be considered adherent.

Characteristics of Health Domains

As displayed in Table 1, the included studies evaluating specific mHealth apps were categorized into the following health domains based on the individual app intervention focus: NCD self-management (17/97, 18%); mental health (20/97, 21%);

XSL•FO
RenderX

substance use (9/97, 9%); nutrition (7/97, 7%); physical activity (6/97, 6%); weight loss (9/97, 9%); multicomponent lifestyle interventions (8/97, 8%); and mindfulness, including breathing and meditation interventions (9/97, 9%). Studies categorized in the domains of NCD self-management, mental health, and substance use were further subcategorized to report intervention and patient-related factors influencing intervention adherence at a more granular level. The studies (17/97, 18%) targeting NCD self-management were subcategorized into diabetes management (6/17, 35%), cancer management (5/17, 29%), respiratory disease management (3/17, 18%), CVD management (2/17, 12%), and medication adherence (1/17, 6%). Studies categorized in the mental health domain (20/97, 21%) were further divided into apps focusing on anxiety (2/20, 10%), depression (9/20, 45%), and multidisciplinary and other (9/20, 45%). The latter subdomain included other mental health problems such as bipolar disorders or combinations of various mental health problems. Studies in the substance use domain were further separated into apps addressing alcohol (2/9, 22%), tobacco (6/9, 67%), or a mix of various substances (1/9, 11%). Another 12% (12/97) of studies, which was a heterogeneous group targeting NCDs other than diabetes, cancer, CVD, respiratory disease, or medication adherence, were clustered into other NCDs (eg, intestinal and renal disease, insomnia, pain, venous leg ulcers, and dyslipidemia).

As outlined in Table 2, the mean number of participants was highest for studies that focused on substance use (2337.6, SD 6344.8; range 9-19,233) and lowest for other NCDs (54.8, SD 48.6; range 15-189), followed by weight loss interventions (73.2, SD 54.4; range 17-176).

As displayed in Table 3, the mean participant age was highest in apps targeting NCD self-management (57.7, SD 7.3; range 45-70.9 years) and lowest for mental health apps (35.9, SD 5.9; range 19.9-46.5 years).

Female populations were generally overrepresented (Table 4), especially in studies conducted on mindfulness interventions (76.7%, SD 20.2%; range 44.8%-100%). Only studies conducted on apps targeting substance use featured more men than women (percentage of women: mean 49.3%, SD 14.1%; range 27.7%-78%).

As outlined in Table 5, studies conducted on apps for weight loss had the most prolonged mean intervention duration (214, SD 216.3; range 65-730 days), and studies on nutrition had the shortest mean intervention duration (52.5, SD 55.7; range 7-172 days).

Table 6 shows distributions of total intended interactions with the apps over the course of the individual studies. The mean number of total intended interactions was highest for apps targeting weight loss (210.5, SD 213.3; range 52-730) and lowest for nutrition apps (31.8, SD 27.8; range 4-82.5).

The distribution of adherence scores by health domain is summarized in Table 7. The mean adherence scores were highest in the domain of other NCDs (69.9%, SD 18.5%; range 33.3%-90.5%), followed by multicomponent lifestyle interventions aimed at changing multiple behaviors simultaneously (61.3%, SD 22.5%; range 32.4%-96%). Apps from the substance use domain had the lowest adherence scores (46.1%, SD 33%; range 9.1%-84%).

Multimedia Appendix 4 [20,26-123] lists the identified intervention-related and patient-related factors with a positive or negative influence on adherence for each health domain in detail. The results per health domain are summarized in the following sections.

XSL•FO
RenderX

Table 1. Included studies evaluating specific mobile health apps categorized by health domain (N=97).

Health domains	Studies, n (%)	References
NCD[a] self-management	17 (18)	[26-42]
Diabetes management	6 (6)	[29-34]
Cancer management	5 (5)	[37-41]
Respiratory disease management	3 (3)	[26-28]
Cardiovascular disease management	2 (2)	[35,36]
Medication adherence	1 (1)	[42]
Mental health	20 (21)	[43-62]
Anxiety	2 (2)	[43,44]
Depression	9 (9)	[45-53]
Multidisciplinary and others	9 (9)	[54-62]
Substance use	9 (9)	[63-71]
Alcohol	2 (2)	[63,64]
Tobacco	6 (6)	[65-70]
Multidisciplinary and others	1 (1)	[71]
Nutrition	7 (7)	[72-77,123]
Physical activity	6 (6)	[78-83]
Weight loss	9 (9)	[84-92]
Multicomponent lifestyle interventions	8 (8)	[93-100]
Mindfulness (including breathing and meditation)	9 (9)	[101-109]
Other NCDs	12 (12)	[110-121]
All domains	97 (100)	[26-121,123]

[a]NCD: noncommunicable disease.

Table 2. Number of participants by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Participants, N	Values, mean (SD)	Values, median (IQR)	Values, range
NCD[a] self-management	15,111	888.9 (2433.3)	56 (113.5-31)	10-9051
Mental health	5710	285.5 (470.1)	81 (231.8-31)	14-1709
Substance use	21,038	2337.6 (6344.8)	99 (683.0-24)	9-19,233
Nutrition	13,042	2173.7 (5195.1)	22 (3342-12)	12-12,777
Physical activity	946	157.7 (147.5)	151 (301-22)	19-301
Weight loss	659	73.2 (54.4)	50 (120.5-28.5)	17-176
MLI[b]	2274	284.3 (531.9)	64.5 (331.3-29.3)	20-1561
Mindfulness	12,608	1400.9 (4031.5)	46 (128-21.5)	15-12,151
Other NCDs	658	54.8 (48.6)	44 (59.8-20.3)	15-189
All domains	72,046	750.5 (2800.7)	56 (129.5-26)	9-19,233

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

XSL•FO
RenderX

Table 3. Age (years) of participants by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Values, mean (SD)	Values, median (IQR)	Values, range
NCD[a] self-management	57.7 (7.3)	56.5 (64.2-52.6)	45.0-70.9
Mental health	35.9 (5.9)	36.6 (40.2-33.9)	19.9-46.5
Substance use	40.8 (9.6)	44.0 (48.8-35.3)	20.5-49.9
Nutrition	44.0 (16.6)	45.0 (60.0-27.2)	22.0-64.7
Physical activity	47.1 (15.2)	42.0 (63.6-38.0)	26.8-68.0
Weight loss	42.5 (11.0)	45.8 (49.6-35.2)	20.0-54.4
MLI[b]	43.7 (18.7)	39.0 (48.8-34.9)	23.6-86.0
Mindfulness	43.7 (14.7)	42.8 (52.8-33.5)	20.2-70.9
Other NCDs	45.6 (10.3)	43.6 (55.2-36.0)	34.0-64.9
All domains	44.6 (12.9)	42.9 (52.7-35.8)	19.9-86.0

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

Table 4. Percentage of women by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Values (%), mean (SD)	Values (%), median (IQR)	Values (%), range
NCD[a] self-management	53.6 (30.9)	50.3 (82.8-31.5)	0-100
Mental health	64.0 (16.5)	65.9 (72.6-58.4)	27.0-95.2
Substance use	49.3 (14.1)	50.5 (55.3-39.6)	27.7-78.0
Nutrition	67.7 (25.5)	71.5 (90.5-44.4)	31.0-94.0
Physical activity	60.6 (16.5)	64.0 (73.8-50.9)	30.4-73.9
Weight loss	60.7 (27.5)	68.9 (81.0-42.5)	1.3-85.0
MLI[b]	66.8 (12.5)	62.9 (78.6-60.0)	51.0-88.3
Mindfulness	76.7 (20.2)	80.2 (94.6-55.7)	44.8-100
Other NCDs	64.0 (26.4)	68.1 (87.5-45.8)	19.0-100
All domains	62.0 (22.8)	63.2 (78.9-48.0)	0-100

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

Table 5. Intervention duration (days) by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Values, mean (SD)	Values, median (IQR)	Values, range
NCD[a] self-management	129.1 (87.9)	91.3 (180-73.5)	30-365
Mental health	97.6 (137)	56 (78.8-30)	28-577.9
Substance use	109.2 (192.1)	56 (90.5-17.6)	14-615
Nutrition	52.5 (55.7)	28 (56-21)	7-172
Physical activity	74.3 (28)	81 (100-49)	28-100
Weight loss	214 (216.3)	180 (273.8-70)	56-730
MLI[b]	79.6 (50.7)	87 (91.3-33.3)	21-182.5
Mindfulness	107.1 (142.3)	42 (202.7-29)	21-365
Other NCDs	111 (126.6)	56 (159.4-31.5)	14-365
All domains	111.4 (132)	60.8 (100-30)	7-730

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

XSL•FO
RenderX

Table 6. Number of intended app interactions by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Values, mean (SD)	Values, median (IQR)	Values, range
NCD[a] self-management	141.6 (184.3)	90 (178.5-35)	6-615.4
Mental health	64.5 (124.6)	39.5 (57.5-6)	4-577.9
Substance use	51.3 (56.2)	19.1 (88.9-9)	4-170.7
Nutrition	31.8 (27.8)	21 (56-10)	4-82.5
Physical activity	60.8 (35.3)	67 (93.3-24)	12-100
Weight loss	210.5 (213.3)	168 (259.3-78)	52-730
MLI[b]	40.2 (36.3)	25.7 (81-6.8)	1-90
Mindfulness	144.1 (243.8)	30 (197.2-24.5)	21-730
Other NCDs	39.8 (35.3)	35 (65.5-5.8)	2-107.1
All domains	90.1 (145.9)	51 (90-20.1)	1-730

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

Table 7. Adherence scores by health domain in the included studies evaluating specific mobile health apps (N=97).

Health domains	Values (%), mean (SD)	Values (%), median (IQR)	Values (%), range
NCD[a] self-management	53.4 (24.7)	63.4 (72.5-25.5)	14.0-89.5
Mental health	56.6 (26.2)	61.0 (83.4-32.2)	15.0-91.5
Substance use	46.1 (33.0)	24.2 (81.1-18.0)	9.1-84.0
Nutrition	49.1 (32.1)	48.9 (84.3-26.8)	2.6-91.4
Physical activity	54.7 (16.6)	55.5 (71.3-39.9)	31.2-72.0
Weight loss	49.1 (21.5)	43.1 (61.6-38.2)	16.0-93.2
MLI[b]	61.3 (22.5)	56.1 (81.6-42.4)	32.4-96.0
Mindfulness	59.0 (18.5)	66.7 (73.7-37.2)	33.3-81.9
Other NCDs	69.9 (18.5)	72.4 (87.6-53.4)	33.3-90.5
All domains	56.0 (24.4)	60.4 (76.0-34.5)	2.6-96.0

[a]NCD: noncommunicable disease.

[b]MLI: multicomponent lifestyle intervention.

NCD Self-management

Factors Influencing Adherence to NCD Self-management Apps

Diabetes Management

Intervention-related factors that positively influenced adherence to diabetes apps included automated and passive data collection within the app [31,34], customized reminders [31], game-based elements [32], and human-like app characteristics [32]. Manual data collection by users [31,34], lack of adjustment to users' personal needs [34], and fast uptake of app activities after initiation [34] were associated with higher intervention dropouts and lower adherence.

Patient-related factors positively affecting adherence to diabetes apps were the following user characteristics: low extraversion [33], high educational level [33], openness to new experiences [33], exacerbated history of diabetes [33,34], and recent diagnosis of the disease [33,34]. Regarding the influence of user age on adherence, the results were contradictory. In one of the studies, users of older age were more adherent [34]; in contrast, another study found that older age was associated with weaker technology acceptance [33].

Cancer Management

Intervention-related factors that positively affected adherence to apps targeting patients with cancer included ongoing contact or telecoaching with health care professionals [37], personalization of users' needs and cultural tailoring [38,39], and customizable reminders and notifications [39,40]. Furthermore, one of the studies showed that a continuous, nondelayed study course positively affected adherence [40].

The following user characteristics were positively associated with adherence: increased age, higher level of education, being married or in a relationship, and higher self-efficacy [37,40]. Furthermore, the active employment status of the users, leading to less available time for the intervention, was associated with lower adherence, especially among female users [38,40].

XSL•FO
RenderX

Respiratory Disease Management

In the context of intervention-related factors, personalization [27], app design, and ease of use [27,28], as well as personal contact or communication with a health care professional, were all positively associated with adherence [27,28].

Patient-related factors included the recruitment strategy and recruitment location. It was shown that users recruited personally and on site were more adherent than users recruited on the web or via social media [26]. Furthermore, the perceived health benefits and the sense of contributing to the science of the users were associated with better adherence [26]. The following user characteristics were negatively related to adherence: higher BMI [27], depression diagnosis [27], low educational level [27], and low smartphone literacy [28].

CVD Management

User interactions with the app through gamification, primarily through a personalized feedback and reward system, were associated with better adherence to apps targeting CVD self-management [36]. In addition, easy communication and data exchange between users and their health coaches positively affected adherence [35].

Patient-related factors that positively affected adherence were the user characteristics of hypertension diagnosis [122] and the high clinical demand of the patient [36]. Lack of technical experience with mobile devices and advanced age of patients were associated with lower adherence [35].

Medication Adherence

One of the studies using an app targeting medication adherence related to various NCDs did not include information about dedicated factors influencing adherence [42].

Quantitative Analysis of NCD Self-management Apps

Across the 5 considered subdomains, the mean adherence score was highest for one app targeting medication adherence (89.5%), followed by the subdomains cancer management (61.4%, SD 19.6%), respiratory disease management (52.7%, SD 21.5%), diabetes management (44.9%, SD 28.6%), and CVD management (41.5%, SD 28.9%). On average, adherence to mHealth apps targeting NCD self-management was 53.4% (SD 24.7%) across all the 5 health domains.

There was a strong positive, significant correlation between adherence score and the average age of study participants (r=0.624; 16/97, 16%; P=.01), which is consistent with the results of the 2 included publications [34,37]. There was no significant correlation between the level of personal support (r_{s14}=0.150; P=.58) or gender (r=0.037; 16/97, 16%; P=.89) and adherence scores across studies targeting NCD self-management.

Mental Health

Factors Influencing Adherence to Mental Health Apps

Anxiety

Compared with manual data collection by users, passive data collection was identified as an effective intervention-related factor in improving adherence in one of the studies [43]. It was also reported that technical problems negatively affected adherence and that iOS users had lower adherence than Android users [43].

Depression

The following intervention-related factors were positively linked with adherence to mHealth apps targeting depression: alternating intervention components and immediate feedback to maintain participant attention [49,52]; individualized features such as personalized representation of intervention progress, encouragement, and daily health tips [47]; offline app functionality and data plan independence [50]; a user-friendly and visually appealing app layout (eg, using a large font or highlighting essential app elements on the home screen) [53]; and evidence-based problem-solving therapies and content [47]. In contrast, a long study duration [49], competitive effects from other apps [50], and declining interest because of waiting times [47] were negatively linked to adherence.

The following user characteristics had a positive impact on adherence: local recruitment [45], ethnic minority background [48], and female gender [49]. In contrast, other characteristics negatively influenced adherence, including Latin America as a geographical origin [47], privacy concerns [47], low income [47], poor baseline depression [48,50] or anxiety, married relationship status [48], and lack of time [53]. In addition, remote recruitment (eg, via a web-based form [45,51]) was identified as a patient-related factor that negatively influences adherence.

Multidisciplinary and Others

The included studies identified individual functions that had a positive impact on adherence, such as crisis plans [58]; self-monitoring and visualization features [56]; tracking of stressful events [58]; tracking mood states with interactive mood charts [58]; visual feedback with personalized graphics interchange format images [62]; and dashboards with information on activity, sleep quality, mood development, and heart rate [61]. In addition, reminders through customizable push notifications were associated with better adherence [62]. Furthermore, the integration of health care professionals was positively linked to adherence [57]. Another study showed that integrating multiple intervention components and avoiding repetition and monotony had a positive impact on adherence [58]. Finally, lack of time for implementation and technical problems were negatively linked to adherence [58].

Patient-related factors positively influencing adherence included the following user characteristics: high IQ [56], increased age [58], increased risk of suicide [60], general interest in the app [54], and a trusting relationship between the person being treated and the organization providing the intervention [55]. In contrast, the following user characteristics negatively affected adherence: long treatment history [54], critical pre-existing conditions (eg, chronic psychotic illness [54]), increased overall mental health burden [56], increased mania-like symptoms [56], privacy concerns [57], or a perceived lack of usefulness of the app [58]. In addition, it was found that the app may lead to an unwanted reminder of one's condition, which negatively affects adherence [58].

XSL•FO
RenderX

Quantitative Analysis of Mental Health Apps

The mean adherence score for the mental health apps was 56.6% (SD 26.2%). The mean adherence scores for apps offering support for anxiety, depression, and other mental health conditions were 67.4% (SD 33.8%), 45.3% (SD 28.1%), and 65.6% (SD 20.93%), respectively.

Regarding the positive effect of incorporating personal support from health care professionals into the intervention reported by Steare et al [57], the correlation between adherence score and the level of personal support for mental health apps, in general, was nonsignificant (r_{s17}=0.230; *P*=.33). Compared with the qualitative data synthesis regarding the relationship between adherence score and average age [58], as well as the gender of student participants [49], no significant relationships were found quantitatively (*r*=0.096; 19/97, 20%; *P*=.70, and *r*=−0.149; 20/97, 21%; *P*=.53, respectively). Regarding the negative effect of long study durations [49], the correlation between adherence score and intervention duration in days was nonsignificant (*r*=−0.127; 20/97, 21%; *P*=.60). Compared with the difference between the iOS and Android operating systems mentioned in the qualitative analysis [43], the differences between the adherence score and smartphone operating system (r_{s17}=0.450; *P*=.05) were likewise positive in the quantitative analysis.

Substance Use

Factors Influencing Adherence to Substance Use Apps

Alcohol

Reminders in the form of daily push notifications were associated with better adherence [63]. Furthermore, personalization and customized content and features were positively linked to adherence [64]. A study also showed that gamification and gamified elements such as levels or rewards positively affected adherence [64]. Finally, variations and options in app design and offer within the app positively affected adherence [64].

The following user characteristics positively affected adherence as patient-related factors: female gender, low-risk alcohol consumption, high education level, reduced substance use, and increased age [64]. Doubts about efficacy and forgetfulness had a negative influence [64].

Tobacco

It was found that reminders in the form of daily push notifications positively affected adherence [69]. In addition, personalization and customized content in the app had the same impact [70]. The integration of and interaction with human coaches were positively associated with adherence [69]. Furthermore, the included studies found some specific features that increased adherence: tracking functions for self-monitoring (eg, as a diary [68,69]), a craving toolbox [69], all-general advice on quitting, and functions for stress and mood management [70].

Regarding patient-related factors, adherence was positively influenced by the following user characteristics: lower initial acceptance of cravings [67], younger age [66], and minimum level of digital skills among users [66].

Multidisciplinary and Others

One of the studies showed that the inclusion of several feedback modules is an effective technique for increasing adherence [71]. Otherwise, the included studies did not provide further information on factors influencing adherence [71].

Quantitative Analysis of Substance Use Apps

On average, the mHealth apps for substance use had an adherence score of 46.1% (SD 33.0%). Apps targeting alcohol use had a higher adherence score (51.5%, SD 38.6%) than those targeting tobacco use (38.0%, SD 32.7%). An app that combined both health behaviors had an adherence score of 83.4%.

Regarding the positive effect of incorporating human coaching into the intervention reported by Webb et al [69], the correlation between adherence score and the level of personal support for substance use apps, in general, was not significant (r_{s6}=0.126; *P*=.77). Compared with the qualitative data synthesis in terms of the relationship between adherence score and average age [64,66] and gender of study participants [64], no significant relationships were found quantitatively (*r*=−0.094, 9/97, 9%, *P*=.81 and *r*=0.394, 9/97, 9%, *P*=.30, respectively).

Nutrition

Factors Influencing Adherence to Nutrition Apps

Personalization of the intervention and certain app functions (personalized overview features of daily goals, recipe suggestions, lookup sections, camera or photograph-taking functions, and barcode scanners) were associated with better adherence to apps targeting nutrition [73,74,76]. Moreover, customized reminders and notifications and the integration of gamification elements combined with incentives enhanced engagement [73]. App handling and user-friendliness further positively influenced adherence [73,76]. The included studies pointed out the importance of the onboarding process, whereas enrollment methods with personal contact [123] had a positive impact, as well as appropriate guidance and tutorials at the beginning [73]. In addition, a relationship between uptake of the intervention activities and adherence was found, whereas starting the intervention on mornings and weekdays, in contrast to weekends, had a positive effect on the use of the mHealth app [75]. Finally, technical difficulties negatively affected adherence [72,73].

Several user characteristics had a positive influence. These were employment at a university [77], female gender [74], high degree of dietary preferences [75], and time and cognitive capacity devoted to the app [75]. The results were inconsistent concerning the age of the user. One of the studies associated older age with more adherence [74], whereas another showed the opposite [73].

Quantitative Analysis of Nutrition Apps

On average, the mHealth apps for nutrition had a mean adherence score of 49.1% (SD 32.1%). The positive effect of the inclusion of personal communication with health care professionals [73,75,123] was confirmed quantitatively, and the correlation between the adherence score and the level of personal support during the study period was strongly positive

XSL•FO
RenderX

and significant (r_{s4}=0.878; *P*=.02). Regarding the relationship between adherence score and age [73,74] and gender of study participants [74], no significant relationships were found quantitatively (*r*=−0.143; 7/97, 7%; *P*=.79, and *r*=0.234; 6/97, 6%; *P*=.66, respectively).

Physical Activity

Factors Influencing Adherence to Physical Activity Apps

Of the 6 included studies targeting physical activity, 2 (33%) showed that customizable push notifications positively affected adherence [79,83]. In addition, the intervention-related factor, gamification, was associated with higher engagement [79]. Furthermore, social features, such as competitions, social comparison, and challenges, positively affected adherence [80,83]. In addition, personalization and customization were positively linked to adherence, especially customizable app functions regarding exercise plans, nutrition suggestions, and calorie lists [79,80,83]. Personal communication with and integration of health care professionals positively affected adherence [82]. Finally, technical difficulties negatively affected adherence [78].

The following user characteristics positively affected adherence: age [79], healthy BMI [79], and a positive attitude toward technology [78]. In contrast, users with increased disease severity, depressive symptoms, low quality of life, and poor access to transportation showed worse adherence [78]. In addition, privacy concerns and a lack of perceived benefits negatively influenced adherence [82].

Quantitative Analysis of Physical Activity Apps

On average, mHealth apps for improving physical activity had an adherence score of 54.7% (SD 16.6%). Regarding the positive effect of personal communication with health care professionals in the intervention [82], the correlation between the adherence score and level of personal support during the study period was not significant (r_{s3}=0.289; *P*=.64). Compared with the qualitative data synthesis regarding the relationship between adherence score and average age [79], no significant correlations were found quantitatively (*r*=0.047; 6/97, 6%; *P*=.93).

Weight Loss

Factors Influencing Adherence to Weight Loss Apps

The included studies focusing on weight loss apps identified a positive influence of reminders in the form of push notifications on adherence [88,92]. Just-in-time intervention components were associated with better adherence [87]. The same was found for newsfeeds with social components [88]. The studies highlighted that personal contact and integration of health care professionals positively influenced adherence [84,85,91]. Moreover, a correlation was found between high adherence and unlimited digital access via the app, as well as providing a data plan with no supplementary costs [89].

Several studies further identified the following user characteristics as patient-related factors to be positively linked to adherence: rural population [85], positive expectations regarding the intervention [88], prior experience with mHealth apps [88], a high sense of responsibility [86], and reinforcements through personal environment [85]. In contrast, dislike of the study equipment [86] and depression symptoms [89] adversely affected adherence.

Quantitative Analysis of Weight Loss Apps

On average, the mHealth apps for weight loss had an adherence score of 49.1% (SD 21.5%).

Regarding the positive effect of incorporating personal communication with health care professionals into the intervention [84,85,91], the quantitative analysis did not reveal a significant correlation (r_{s7}=0.174; *P*=.65).

Mindfulness

Factors Influencing Adherence to Mindfulness Apps

As most mindfulness apps that met the inclusion criteria of this review did not distinctively aim to treat a chronic mental condition but rather to increase well-being and reduce work-related stress, we categorized them as a separate category. Factors identified in the mindfulness domain may still be relevant for the mental health domain, as mindfulness-based therapy has been cited as a promising intervention for treating anxiety and depression in previous research [124].

The included studies reported that automated and interactive data collection and processing positively affected adherence [101,106]. The studies also showed that customizable features such as tracking stress and mood [109], visualizing personal progress, and immediate feedback positively affected adherence [103]. In addition, using in-app tutorials or video content was associated with better adherence [103]. Furthermore, time and place influenced adherence: users who used the app in the evening and at home were more adherent in the long term [106,107]. In contrast, extensive app interactions and lack of variety in the app content harmed adherence [107].

Regarding patient-related factors, the studies identified the following user characteristics to be positively related to adherence: increased age [102,108], positive expectations toward the app [106,108], intrinsic motivation [108], and a current physical diagnosis [106] in contrast to a mental health diagnosis.

Quantitative Analysis of Mindfulness Apps

On average, mHealth apps related to mindfulness had an adherence score of 59.0% (SD 18.5%). Regarding the positive relationship between adherence score and average age [102,108], the quantitative analysis also yielded a moderately positive but nonsignificant correlation (*r*=0.404; 9/97, 9%; *P*=.28).

Multicomponent Lifestyle Interventions

Factors Influencing Adherence to Multicomponent Lifestyle Intervention Apps

Approximately 8% (8/97) of studies targeted mHealth apps focusing on multiple lifestyle behaviors (mostly a combination of physical activity, diet, weight loss, and sometimes sleep, stress, or headaches).

The included studies reported that the integration of health care professionals during the intervention, app usability, and language positively influenced adherence [95,98]. However, it was also

XSL•FO
RenderX

shown that social networking and competition through social comparison in terms of physical activity level had a positive impact only if individuals had a healthy BMI [97]. In addition, app features such as audiovisual presentation of health-related information or reminders in the form of push notifications positively affected adherence [93,98]. Finally, personalization and tailoring of the app to customized needs (eg, through gamification) had a positive impact on adherence [97,98].

In addition to these factors, the following characteristics also positively affected adherence: increased age [95,100] and trust in the health care professionals of the intervention [95,100]. Finally, other characteristics negatively affected adherence. These included the lack of engagement of other participants in social comparison features [97], negative emotions related to self-monitoring during periods of weight [97], shift work schedules [98], and technical difficulties in using the app [93].

Quantitative Analysis of Multicomponent Lifestyle Intervention Apps

On average, mHealth apps targeting multicomponent lifestyle interventions had an adherence score of 61.3% (SD 22.5%).

The positive effect of integrating health care professionals as personal support into the intervention [95,98] could not be analyzed quantitatively as none of the included multicomponent lifestyle interventions offered consistent, continuous support by health care professionals (only during the onboarding phase). Regarding the mentioned positive relationship between adherence score and average age [95,100], the quantitative analysis also yielded a moderately positive but nonsignificant correlation (r=0.416; 8/97, 8%; P=.31).

Other NCDs

Factors Influencing Adherence to Other NCD Apps

In a study on an mHealth app treating insomnia, better adherence was linked to ease of use and the easiness of therapy directives [110]. Another study on insomnia treatment found that easy access and reminder options or notifications had a positive impact [113]. In the field of chronic pain management and interventions, a study found a positive impact on adherence to microinteractions [111]. It was also shown in the same field that personalization had a positive impact [112]. In a study on the care of advanced chronic kidney disease, the integration of complementary visits to health care professionals or to a clinic showed a positive impact on adherence [117]. Furthermore, blood pressure and test result features and an automatic integrated transfer of blood pressure readings had a positive impact [117].

Regarding patient-related factors, a study on the treatment of irritable bowel syndrome showed that the simultaneous use of other technical devices positively affected adherence. Another study showed that high anxiety scores negatively affected adherence [119]. A study on the effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders found that older age had a positive impact on adherence. In contrast, low acceptability, lack of time, health problems, and lack of motivation had a negative impact [118]. The user characteristic of the female gender positively influenced adherence in a study about an app for a lower leg physical activity intervention for individuals with chronic venous leg ulcers [114]. Furthermore, in a study on an mHealth app targeting inflammatory bowel disease, it was found that old age, low level of education, and lack of perceived usefulness negatively affected adherence [115].

Quantitative Analysis of Other NCD Apps

On average, mHealth apps categorized in the domain of *other NCDs* had an adherence score of 69.9% (SD 18.5%). Regarding the positive effect of complementary visits to health care professionals for the care of chronic kidney disease [117], the correlation between the adherence score and the level of personal support during the study period was nonsignificant (r_{s10}=0.290; P=.36). The quantitative analysis supports the finding that female participants are more adherent [118] to some degree, as the correlation between the adherence score and mean percentage of female participants was moderately positive but nonsignificant (r=0.385; 12/97, 12%; P=.22). Regarding the relationship between adherence score and average age [118], the quantitative analysis yielded conflicting results (r=−0.619; 12/97, 12%; P=.03).

Multi-Domain Review

One of the two included systematic reviews featured a multi-domain review focusing on uptake and engagement with mHealth apps in various health domains [20]. First, it showed a positive impact of goal setting, reward offerings, complementary web access, coping games, and self-monitoring. In addition, the low cost of the app helped increase acceptance [20].

The external influence of using an app through health care professionals, friends, and family or by reading user reviews was outlined as having a positive influence. Furthermore, community networking and the connection between the app and health professional support had a positive influence [20]. The study also found the following user characteristics to be positively linked to adherence: female gender, aged <44 years, living in urban areas, good educational level, high income, curiosity, higher health literacy, and app awareness [20]. In addition, interactivity, an established routine to use the app, and customization of the app had a positive impact [20]. In contrast, cognitive overload and unmet expectations negatively influenced adherence [20].

Explorative Analysis of Adherence Scores

To gain further insights into the universal applicability of the results identified in individual health domains, a quantitative analysis was conducted on the total number of primary studies included. The analysis revealed a positive correlation between adherence score and level of personal support during the study period (r_{s91}=0.199; P=.06). With respect to various user characteristics, the quantitative analysis did not find significant differences in either average age (r=0.105; 94/97, 97%; P=.32) or gender distribution (r=−0.031; 95/97, 97%; P=.77). Furthermore, no significant quantitative differences were found between healthy participants and participants with chronic diseases (r_{s95}=−0.049; P=.63). A quantitative comparison of studies with monetary incentives and those without such

XSL•FO
RenderX

incentives also revealed no significant effect on adherence scores (r_{s92}=0.000; *P*=.99). However, the monetary value of the incentive, measured in US dollars, had a significant effect on the adherence score (*r*=0.465; 30/97, 31%; *P*=.01). Apps that were only offered in the context of scientific studies had a significantly higher adherence score than those that were publicly available via app stores (r_{s95}=0.324; *P*=.001). The quantitative analysis did not find any significant differences between Android and iOS with regard to the adherence score (r_{s83}=0.019; *P*=.87). Furthermore, the quantitative analysis showed a higher adherence score for apps developed by private app development companies than for those developed by public institutions or research groups (r_{s95}=0.164; *P*=.11). The correlation between adherence score and intervention duration (*r*=−0.138; 97/97, 100%; *P*=.18) was negative but positive in relation to the number of intended app interactions per day (*r*=0.176; 97/97, 100%; *P*=.09). The comparison of adherence scores and the total number of intended app interactions (*r*=0.040, 97/97, 100%; *P*=.70) yielded no significant results. Studies with a higher number of app users had significantly lower adherence scores (*r*=−0.228; 96/97, 99%; *P*=.03).

Discussion

Intervention-Related Factors Influencing Adherence

Regarding the first research question, the intervention-related factors described in the following sections were identified most frequently across all health domains, suggesting universal applicability to increase mHealth app adherence relative to the intended use.

User-friendliness and Technical Stability

Approximately 18% (17/97) of studies from 6 health domains cited a user-friendly app design or technical stability as criteria for increased app use [27,28,43,50,53,58,72-74,76,78,93-95,110,113,115,116]. The term *user-friendliness* describes a software interface that enables a simple, clean, intuitive, and reliable user experience (UX). App developers can thus promote adherence by making the app easy to use and providing a compelling and visually appealing app design (eg, by using sufficiently large fonts or highlighting essential app elements) [53]. Technical problems can be reduced through closed beta tests while optimizing the UX through user interface or UX design changes before app release. Accordingly, the quantitative analysis revealed higher adherence scores for apps created by private app development companies, which may have more technical expertise in developing and publishing apps than public institutions or research groups. Quantitative analysis also revealed higher adherence scores for apps developed by private companies, which may have more expertise across the value chain than public institutions or research groups. As most of the included mHealth apps had a pilot character (ie, developed by small academic teams with no or only short testing periods), it can be assumed that the current body of mHealth apps does not yet realize its full potential in terms of usability and technical stability, thus indicating the potential to improve adherence.

Personalization, Customization, and Tailoring

Approximately 16% (16/97) of studies from 8 different health domains reported a positive impact of personalized content on adherence [27,38,39,58,62,64,66,73,74,76,83,85,91,97,98,113]. This included individualized app features (such as a crisis plan), metrics, visualizations based on individual user data (eg, displaying intervention progress), personalized feedback and health suggestions, and individualized app content tailored to the needs and characteristics of users. These findings align with previous reviews that have summarized that an individualized app positively influences user engagement [20,125]. Accordingly, developers of mHealth apps should consider the target group's characteristics and needs in the app design process and ideally make the app tailored to a specific user group, personalized to the individual, and customizable.

Individualized Reminders

In 13% (13/97) of studies across 8 different health domains, reminders, primarily realized through push notifications, were highlighted as an effective method of improving adherence to mHealth app interventions [31,39,40,62,63,69,73,79,83,88,92-94,113]. Essential to the success of this technique is the consideration of users' individual needs in terms of their schedule, as a user's lack of time undermines adherence [38-40,53,58,118]. Ideally, users receive reminders when they are in a state of receptivity, which has also been suggested by previous research on just-in-time-adaptive interventions [126,127]. In this regard, working and leisure time schedules should be considered, and the timing of reminders should be adapted accordingly, particularly when patients are in the privacy and comfort of their homes. A recent review reported that reminders are helpful for people with busy schedules and when they are forgetful but also mentioned the risk that push notifications can threaten users' social identity if they are received at an inappropriate time or place [20]. Therefore, users should be able to customize the reminders and adapt them to their circumstances.

Personal Support From Health Care Professionals

In 12% (12/97) of studies from 8 different health domains, personal support from a health care professional during the intervention was cited as a reason for improved adherence [28,35,37,57,69,82,84,91,95,98,117,123]. In this context, the integration of health care professionals past the initial app onboarding can be realized in various ways, including regular clinic visits, complementary telephone support, and communication options with health care professionals integrated into the app. In addition, apps can facilitate communication between patients and health care professionals through automated data exchanges. Quantitative analysis revealed a positive correlation between adherence and level of personal support during the study period for all health domains, which was also confirmed by previous studies [20,128]. Consequently, it can be assumed that hybrid systems that combine automated app content with elements of human support achieve higher adherence rates than those achieved by interventions without human support. Although the ideal ratio between human-computer interactions and sole human interactions in mHealth app interventions remains to be explored, new

XSL•FO
RenderX

technologies such as conversational agents show promising results in simulating personal support without the need for human support and may enable increased levels of automation [129-132].

Elements of Gamification and Social Features

Elements of gamification were described as effective in increasing adherence in 12% (12/97) of studies across 6 different health domains [32,36,47,49,52,64,73,78-80,83,97,98]. These elements included levels, reward systems, social characters, contests, and leaderboards. This aligns with other reviews that list rewards and games as factors that positively affect adherence [20,125]. However, although game elements such as social competitions may increase engagement by encouraging others, the idea of defeating peers may also have a negative influence [20]. In 6% (6/97) of studies from 5 health domains, social features were found to positively affect adherence; for example, in the form of social networks, contests, leaderboards, or newsfeeds with social components [52,73,75,83,88,97]. However, social components should be included with caution and ideally tested, as social comparison with less-motivated participants can also harm adherence [97]. Accordingly, a recent review concluded that social contests increase engagement but may also have a negative effect [20]. In general, this study supports previous research outlining the positive effects of gamification and social features on DHIs [133-141].

Passive and Automated Data Collection, Processing, and Transmission

The positive influence of passive and automated data collection, processing, and transmission was reported in 5% (5/97) of studies from the health domains NCD self-management [31,34], other NCDs [117], mindfulness [101], and meditation [106]. Thus, developers can increase adherence by automating repetitive tasks to reduce user burden. In this regard, developers are advised to use smartphone sensor technology (eg, a camera to capture food data or accelerometer data to capture physical activity) and complementary devices (eg, a smartwatch to measure heart rate) to remove the repetition and monotony of intervention tasks, which are listed as reasons for nonadherence [58,60,70].

Monetary Incentives

Monetary incentives such as vouchers, lottery tickets, or direct cash contributions were given to participants in 31 of the 97 included primary studies [30,33,44,45,47,48,51,53,56,61,62,64,65,68-72,77,81,88,90,98,101,104-106,108,109,114,116]. However, the qualitative synthesis did not yield results regarding the effect of such monetary incentives as an additional intervention component on adherence. The quantitative comparison between studies with monetary incentives and those without such incentives also found no effect on adherence. However, the monetary value of the incentive, measured in US dollars, had a significantly positive effect on adherence scores. In the context of these findings, app developers may consider whether monetary incentives are helpful and the level of compensation that is sufficient to achieve a relevant effect.

Other Intervention-Related Factors

Other notable intervention-related factors included integrating an app tutorial [73,103], presenting information in audiovisual formats [98], and offering a large variety of app content [64]. Approximately 3% (3/97) of studies also noted financial costs (eg, data plan use) as a barrier to adherence [20,89,92]. Therefore, developers may want to consider offering their mHealth app free of charge and only transferring large amounts of data when the device is connected to a wireless network. Other studies in this review also reported that time delays within the intervention, long intervention durations, low engagement of other participants, and competitive effects of other mHealth apps were associated with low adherence. The included studies also outlined data protection and user privacy as positively affecting adherence [20,47,57,82], which aligns with previous research calling developers to create robust and transparent mHealth apps that satisfy security and privacy demands to foster user acceptance and trust [142-145].

Patient-Related Factors Influencing Adherence

Regarding the second research question, the patient-related factors described in the following section were identified.

Characteristics of Study Participants

Approximately 43% (42/97) of studies from 9 different health domains reported a wide variety of user characteristics that affect adherence, including age, gender, place of residence, marital status, health status, treatment history, education, employment status, income, and work hours [26-28,33-38,40,47-50,54,56,58,60,64,66,67,73-75,77-79,85,86,88,89,95,98,100,102,106,108,114,115,118,119,122]. The quantitative analysis did not reveal significant effects of average age, gender distribution, or pre-existing conditions on adherence across all included apps. Consequently, the results from previous reviews, which indicate higher engagement among female or younger users, could not be replicated [20,125]. The specific health domains of NCD self-management and other NCDs showed a significant correlation between adherence scores and the average age of study participants, with the first one being positive and the latter being negative. Further research is required to understand the effects of sociodemographic characteristics and health status on adherence to mHealth apps. The findings of this study suggest that these effects may vary depending on the targeted health domain.

User characteristics associated with a low adherence were lack of technical competence, lack of health literacy, and lack of experience with mHealth apps, which could potentially be improved through preintervention training. Other negative factors, such as privacy concerns, low expectations of the app, and low trust in the health care professionals conducting the intervention, could potentially be challenged through personal communication in the onboarding phase (eg, by discussing the privacy policy or outlining intervention benefits). As lack of time on the users' side was referenced to negatively affect adherence [53,58,107,118], helping patients with time management might also have a positive effect.

XSL•FO
RenderX

Type of Participant Recruitment

In 4% (4/97) of studies from the fields of NCD self-management [26], mental health [45,51], and nutrition [123], the user recruitment channel was mentioned as a relevant factor affecting adherence. Users who were made aware of the intervention on the web (eg, via social media) had lower adherence than users recruited locally and in person. This could explain why apps that were only offered in the context of studies on personal onboarding processes had significantly higher adherence scores than those publicly accessible via app stores. As the mHealth sector further matures and more mHealth apps are made available to the public via app stores, this factor might have an increasingly negative effect on overall adherence. Developers may overcome this issue by optimizing the UX of their mHealth app in the onboarding phase. As highlighted previously, offering personal support from health care professionals before and during the onboarding process is likely to increase adherence.

Adherence Scores Across Health Domains

Regarding the third research question, this review outlined the differences between health domains in terms of adherence scores. The adherence score of all 97 included mHealth apps averaged 56.0% (SD 24.4%), representing a generally higher adherence level to mHealth apps than previous research suggests [9,10,146]. This could be attributed to the fact that some studies excluded participants who did not perform a certain level of activity (eg, downloading the app). Regarding the short intervention periods with a median of 60.8 days, it is questionable whether the included health apps could reach similar adherence scores in more prolonged studies. Adherence scores by health domain *were highest for the other NCDs (69.9%, SD 18.5%)*, a heterogeneous group of mHealth apps targeting less common NCDs such as intestinal and renal disease, insomnia, pain, venous leg ulcers, and dyslipidemia. Multicomponent lifestyle interventions aimed at changing multiple behaviors simultaneously (61.3%, SD 22.5%) had the second-highest adherence scores, whereas apps targeting substance use had the lowest (46.1%, SD 46.1%). The relatively low adherence levels of mHealth apps treating substance use could be explained by the nature of their intervention design, making it difficult to differentiate between nonadherent users and users who stopped using the app after a successful behavior change. Another explanation could be that substance use disorders are comorbid with depressive disorders. Several qualitative findings indicate that symptoms or a diagnosis of depression negatively affect adherence [27,48,50,78,89]. This also aligns with our findings that apps offering depression support had an even lower average adherence score of 45.3% (SD 28.1%).

Another explanation and potential bias for the difference in adherence scores are asymmetric distributions of trials compared with real-world applications within health domains, represented by the mean number of participants (substance use 2337.6, SD 6344.8; multicomponent lifestyle interventions 284.3, SD 531.9; and other NCDs 54.8, SD 48.6). In general, studies with a higher number of participants had significantly lower adherence scores (r=−0.228; 96/97, 99%; P=.03). Furthermore, apps that were only offered in the context of scientific studies had a significantly higher adherence score than those publicly available via app stores (r_{s95}=0.324; P=.001). Surprisingly, these differences could not be explained by longer study duration (r=−0.138; 97/97, 100%; P=.18) or the number of intended app interactions (r=0.040; 97/97, 100%; P=.70). Thus, our study supports and provides quantitative evidence for previous findings outlining engagement within trials to significantly differ from real-world applications [147-149]. Further real-world longitudinal studies are necessary to explain these differences.

Although calculating an adherence score as the ratio between intended and actual use has several limitations, this exploratory approach enabled the quantification and comparison of adherence across different mHealth apps to a reasonable extent. The quantitative analysis of adherence scores yielded few significant results but fit qualitative findings regarding positive or negative influences on adherence in most instances, which supports the potential validity of the concept.

The results of explorative analysis based on adherence scores should be considered cautiously. Although we did not find a significant correlation between study duration or the number of intended app interactions and the adherence score, it is possible that other factors, such as patient characteristics, could influence adherence. Further research is needed to establish effective adherence measures. By reporting this novel quantified measure of adherence for individual studies and collectively for defined health domains, we hope researchers will test and challenge our approach, potentially developing more effective measures that help us quantify adherence and make it comparable across heterogeneous groups of mHealth apps.

Limitations

This study had several limitations. The first limitation is the heterogeneity of the included studies. The studies differ in terms of the characteristics of the target populations (eg, type of pre-existing condition, age, gender, education level, comorbidities, employment status, and experience with mobile technologies) and study duration (a few weeks to over a year). In addition, mHealth apps within the studies and their intended use varied significantly. The problem of the undiversified reach of mHealth interventions, predominantly including female and White participants living in high-income countries [150], also accounts for this study. Approximately 90% (87/97) of these studies were conducted in North America, Australia, or Europe. Similarly, women were overrepresented, with an average proportion of 62% (SD 22.8%). Most included studies had a pilot character, with 80% (77/96) of the studies including <200 participants and 73% (71/97) of the studies having a duration of <100 days. Whether individual study results can be replicated in more controlled and longitudinal studies in the future is questionable. In addition, there were differences in terms of additional monetary incentives and the level of personal support complementary to the use of mHealth interventions. Moreover, the mHealth apps investigated in the studies showed substantial heterogeneity in their goals (eg, increasing physical activity and reducing tobacco consumption). Overall, this limited the generalizability of our results.

This limitation was overcome by categorizing the results into different health domains. However, mHealth apps also exhibited

XSL•FO
RenderX

key differentiating characteristics within their health domain, such as their stage of development (prototype vs established app), developer (nonprofit vs private company), the use of peripheral devices (eg, a smartwatch for passive data collection), app features (eg, social features such as leaderboards and elements of gamification), or the level of quality (eg, in terms of the user interface, UX, or technical stability). Another limitation is that few studies considered individual intervention components separately and evaluated their effectiveness, which complicates the identification of intervention-related factors. A further limitation of this study is the inclusion of nonrandomized studies, which, on the one hand, allowed a more extensive consideration of objective app use data but, by contrast, precluded conducting a risk of bias assessment.

The calculation of adherence scores as the ratio between intended and actual use also implies some noteworthy drawbacks. The intended use of the mHealth app, which was derived from the study design or study author comments, was also heterogeneous and differed depending on the mHealth app design (eg, tracking daily symptoms or completing a certain number of coaching sessions), interaction frequency (eg, daily or weekly), and interaction duration (weeks to years). Furthermore, the intended use was rarely justified by applying theory, evidence, or rationale, which has also been addressed in previous studies [13]. In some instances, the intended use could be derived from the intervention design; however, many studies had to be excluded as the app's intended use was unclear or not stated at all. However, it can be positively highlighted that the actual mHealth app use extracted for the adherence score was based on objective app use data in approximately 96% (92/96) of cases.

Conclusions

This study contributes to the scarce scientific evidence on factors influencing adherence to mHealth apps and is the first to derive quantified adherence scores for various health domains to validate qualitative findings and explore adherence benchmarks.

This paper contains various detailed presentations of the central results. The most detailed presentation of adherence factors extracted from individual studies is outlined in Multimedia Appendix 4. We further classify the factors within the defined health domains and report the results collectively. Finally, we discuss the most common factors influencing adherence across all the health domains. As mHealth apps within health domains remain heterogeneous, we encourage readers to always consider information from the corresponding individual studies outlined in Multimedia Appendix 4 when implementing their interventions according to the factors reported in this review.

Our findings indicate that the following intervention-related factors positively influence mHealth app adherence: user-friendly and technically stable app design, customizable push notifications, personalized app content, passive data tracking, an integrated app tutorial, offering the app free of charge, and the integration of personal support into the mHealth app. Furthermore, gamification and social features show promising effects but may be limited to specific health domains. Time delays within the intervention, long intervention durations, low engagement of other participants, and the competitive effects of other mHealth apps were associated with low mHealth app adherence.

Regarding patient-related factors, the following user characteristics were associated with low mHealth app adherence: lack of technical competence, low health literacy, low self-efficacy, low education level, mental health burden, lack of experience with mHealth apps, privacy concerns, low expectations of the app, low trust in health care professionals conducting the intervention, and lack of time on the users' side. Age, gender, and pre-existing condition were frequently mentioned factors but differed across and sometimes conflicted within the health domains. Furthermore, personal user recruitment appeared to positively influence adherence as opposed to web-based user recruitment.

The adherence score of the 97 included mHealth apps averaged 56.0% (SD 24.4%). Adherence scores were highest for mHealth apps targeting less common NCDs such as intestinal and renal disease, insomnia, pain, venous leg ulcers, and dyslipidemia. Multicomponent lifestyle interventions had the second-highest adherence scores, whereas apps targeting substance use had the lowest. Exploratory analysis of adherence scores revealed quantitative evidence for higher adherence rates within trials than in real-world applications.

Overall, research on the factors that positively or negatively influence adherence to mHealth apps is still limited. The underlying studies often had a pilot character with a short study duration, and the implementation of techniques was inconsistent. As most mHealth apps contain multiple intervention components, causal statements about individual factors are not possible and require more controlled and longitudinal studies in the future. To facilitate future research on mHealth app adherence; researchers should clearly outline and justify the app's intended use, report objective data on actual use relative to the intended use; and ideally, provide long-term use and retention data. Further research is needed to establish effective adherence measures.

Acknowledgments

This study was partially supported by the Federal Office of Public Health of the Swiss Confederation and CSS Insurance, Switzerland. This research was also partially supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise Program and was partly conducted at the Future Health Technologies program, which was established collaboratively between the Swiss Federal Institute of Technology in Zürich and the National Research Foundation Singapore. This research was supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise Program. The Federal Office of Public

XSL•FO
RenderX

Health, CSS Insurance, and the National Research Foundation had no role in the study design, data collection, data analysis and interpretation, writing the manuscript, or reviewing and approving the manuscript for publication.

Authors' Contributions

RJ conceived the study using inputs from TK. RJ, SH, AMR, and JLM contributed to search strategy, screening, data extraction, and data synthesis. RJ wrote the report and conducted the statistical analyses. TK, SH, EF, JLM, and ASS provided methodological guidance and feedback on the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

RJ, EF, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative between the Department of Management, Technology, and Economics at the Swiss Federal Institute of Technology in Zürich and the Institute of Technology Management at the University of St Gallen, which is partly funded by the Swiss health insurer CSS. The CSS was not involved in any stage of the study. EF and TK are also the cofounders of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in this study.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - jmir_v24i5e35371_app1.docx]

Multimedia Appendix 2
Applied search strategies.
[DOCX File , 29 KB - jmir_v24i5e35371_app2.docx]

Multimedia Appendix 3
Data extraction sheet.
[XLSX File (Microsoft Excel File), 54 KB - jmir_v24i5e35371_app3.xlsx]

Multimedia Appendix 4
Intervention-related and patient-related factors influencing adherence by health domain.
[DOCX File , 49 KB - jmir_v24i5e35371_app4.docx]

References

1. Noncommunicable diseases. World Health Organization. 2021 Apr 13. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2022-04-12]
2. mHealth. eHealth Suisse. 2021 Aug 03. URL: https://www.e-health-suisse.ch/gemeinschaften-umsetzung/ehealth-aktivitaeten/mhealth.html [accessed 2022-04-12]
3. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015 May 19;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
4. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
5. 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 Sep 20;1:48 [FREE Full text] [doi: 10.1038/s41746-018-0059-8] [Medline: 31304327]
6. Aitken M, Clancy B, Nass D. The Growing Value of Digital Health - Evidence and Impact on Human Health and the Healthcare System. IQVIA Institute. 2017 Nov 7. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health [accessed 2022-04-12]
7. Hesser H. Estimating causal effects of internet interventions in the context of nonadherence. Internet Interv 2020 Aug 29;21:100346 [FREE Full text] [doi: 10.1016/j.invent.2020.100346] [Medline: 32983907]
8. 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] [Medline: 15829473]
9. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
10. 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 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]

XSL•FO
RenderX

11. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
12. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
13. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
14. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019 Sep 04;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]
15. Trinh H, Shamekhi A, Kimani E, Bickmore TW. Predicting user engagement in longitudinal interventions with virtual agents. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents. New York, NY, USA: Association for Computing Machinery; 2018 Nov Presented at: IVA '18; November 5-8, 2018; Sydney, Australia p. 9-16 URL: https://doi.org/10.1145/3267851.3267909 [doi: 10.1145/3267851.3267909]
16. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018 Sep 08;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]
17. Kernebeck S, Busse TS, Ehlers JP, Vollmar HC. [Adherence to digital health interventions: definitions, methods, and open questions]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1278-1284 [FREE Full text] [doi: 10.1007/s00103-021-03415-9] [Medline: 34559252]
18. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
19. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012 Mar;36(1):157-178. [doi: 10.2307/41410412]
20. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
21. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
22. Nahum-Shani I, Rabbi M, Yap J, Philyaw-Kotov ML, Klasnja P, Bonar EE, et al. Translating strategies for promoting engagement in mobile health: a proof-of-concept microrandomized trial. Health Psychol 2021 Dec;40(12):974-987 [FREE Full text] [doi: 10.1037/hea0001101] [Medline: 34735165]
23. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol (forthcoming) 2022 Mar 17. [doi: 10.1037/amp0000983] [Medline: 35298199]
24. Teague S, Youssef GJ, Macdonald JA, Sciberras E, Shatte A, Fuller-Tyszkiewicz M, SEED Lifecourse Sciences Theme. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Med Res Methodol 2018 Nov 26;18(1):151 [FREE Full text] [doi: 10.1186/s12874-018-0586-7] [Medline: 30477443]
25. Booker CL, Harding S, Benzeval M. A systematic review of the effect of retention methods in population-based cohort studies. BMC Public Health 2011 Apr 19;11:249 [FREE Full text] [doi: 10.1186/1471-2458-11-249] [Medline: 21504610]
26. Hui CY, McKinstry B, Walton R, Pinnock H. A mixed method observational study of strategies to promote adoption and usage of an application to support asthma self-management. J Innov Health Inform 2019 Jan 09;25(4):243-253 [FREE Full text] [doi: 10.14236/jhi.v25i4.1056] [Medline: 30672405]
27. Rodriguez Hermosa JL, Fuster Gomila A, Puente Maestu L, Amado Diago CA, Callejas González FJ, Malo De Molina Ruiz R, et al. Compliance and utility of a smartphone app for the detection of exacerbations in patients with chronic obstructive pulmonary disease: cohort study. JMIR Mhealth Uhealth 2020 Mar 19;8(3):e15699 [FREE Full text] [doi: 10.2196/15699] [Medline: 32191213]
28. Deng N, Chen J, Liu Y, Wei S, Sheng L, Lu R, et al. Using mobile health technology to deliver a community-based closed-loop management system for chronic obstructive pulmonary disease patients in remote areas of China: development and prospective observational study. JMIR Mhealth Uhealth 2020 Nov 25;8(11):e15978 [FREE Full text] [doi: 10.2196/15978] [Medline: 33237036]
29. Sittig S, Wang J, Iyengar S, Myneni S, Franklin A. Incorporating behavioral trigger messages into a mobile health app for chronic disease management: randomized clinical feasibility trial in diabetes. JMIR Mhealth Uhealth 2020 Mar 16;8(3):e15927 [FREE Full text] [doi: 10.2196/15927] [Medline: 32175908]
30. Dugas M, Crowley K, Gao GG, Xu T, Agarwal R, Kruglanski AW, et al. Individual differences in regulatory mode moderate the effectiveness of a pilot mHealth trial for diabetes management among older veterans. PLoS One 2018 Mar 7;13(3):e0192807 [FREE Full text] [doi: 10.1371/journal.pone.0192807] [Medline: 29513683]

XSL•FO
RenderX

31. Padhye NS, Wang J. Pattern of active and inactive sequences of diabetes self-monitoring in mobile phone and paper diary users. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:7630-7633. [doi: 10.1109/EMBC.2015.7320159] [Medline: 26738059]
32. Gong E, Baptista S, Russell A, Scuffham P, Riddell M, Speight J, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 05;22(11):e20322 [FREE Full text] [doi: 10.2196/20322] [Medline: 33151154]
33. Su J, Dugas M, Guo X, Gao GG. Influence of personality on mHealth use in patients with diabetes: prospective pilot study. JMIR Mhealth Uhealth 2020 Aug 10;8(8):e17709 [FREE Full text] [doi: 10.2196/17709] [Medline: 32773382]
34. Böhm AK, Jensen ML, Sørensen MR, Stargardt T. Real-world evidence of user engagement with mobile health for diabetes management: longitudinal observational study. JMIR Mhealth Uhealth 2020 Nov 06;8(11):e22212 [FREE Full text] [doi: 10.2196/22212] [Medline: 32975198]
35. Harzand A, Witbrodt B, Davis-Watts ML, Alrohaibani A, Goese D, Wenger NK, et al. Feasibility of a smartphone-enabled cardiac rehabilitation program in male veterans with previous clinical evidence of coronary heart disease. Am J Cardiol 2018 Nov 01;122(9):1471-1476 [FREE Full text] [doi: 10.1016/j.amjcard.2018.07.028] [Medline: 30217377]
36. Kaplan AL, Cohen ER, Zimlichman E. Improving patient engagement in self-measured blood pressure monitoring using a mobile health technology. Health Inf Sci Syst 2017 Oct 7;5(1):4 [FREE Full text] [doi: 10.1007/s13755-017-0026-9] [Medline: 29081974]
37. Crafoord MT, Fjell M, Sundberg K, Nilsson M, Langius-Eklöf A. Engagement in an interactive app for symptom self-management during treatment in patients with breast or prostate cancer: mixed methods study. J Med Internet Res 2020 Aug 10;22(8):e17058 [FREE Full text] [doi: 10.2196/17058] [Medline: 32663140]
38. Baik SH, Oswald LB, Buscemi J, Buitrago D, Iacobelli F, Perez-Tamayo A, et al. Patterns of use of smartphone-based interventions among Latina breast cancer survivors: secondary analysis of a pilot randomized controlled trial. JMIR Cancer 2020 Dec 08;6(2):e17538 [FREE Full text] [doi: 10.2196/17538] [Medline: 33289669]
39. Benze G, Nauck F, Alt-Epping B, Gianni G, Bauknecht T, Ettl J, et al. PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. Ann Palliat Med 2019 Apr;8(2):104-111 [FREE Full text] [doi: 10.21037/apm.2017.07.05] [Medline: 29156896]
40. Min YH, Lee JW, Shin YW, Jo MW, Sohn G, Lee JH, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May 23;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
41. Greer JA, Jacobs JM, Pensak N, Nisotel LE, Fishbein JN, MacDonald JJ, et al. Randomized trial of a smartphone mobile app to improve symptoms and adherence to oral therapy for cancer. J Natl Compr Canc Netw 2020 Feb;18(2):133-141. [doi: 10.6004/jnccn.2019.7354] [Medline: 32023526]
42. Mira JJ, Navarro I, Botella F, Borrás F, Nuño-Solinís R, Orozco D, et al. A Spanish pillbox app for elderly patients taking multiple medications: randomized controlled trial. J Med Internet Res 2014 Apr 04;16(4):e99 [FREE Full text] [doi: 10.2196/jmir.3269] [Medline: 24705022]
43. Pham Q, Khatib Y, Stansfeld S, Fox S, Green T. Feasibility and efficacy of an mHealth game for managing anxiety: "Flowy" randomized controlled pilot trial and design evaluation. Games Health J 2016 Feb;5(1):50-67. [doi: 10.1089/g4h.2015.0033] [Medline: 26536488]
44. Miller CB, Gu J, Henry AL, Davis ML, Espie CA, Stott R, et al. Feasibility and efficacy of a digital CBT intervention for symptoms of Generalized Anxiety Disorder: a randomized multiple-baseline study. J Behav Ther Exp Psychiatry 2021 Mar;70:101609. [doi: 10.1016/j.jbtep.2020.101609] [Medline: 32950939]
45. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language Behavioral Activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among united states Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
46. 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 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
47. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
48. 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] [Medline: 27998876]
49. Economides M, Lehrer P, Ranta K, Nazander A, Hilgert O, Raevuori A, et al. Feasibility and efficacy of the addition of heart rate variability biofeedback to a remote digital health intervention for depression. Appl Psychophysiol Biofeedback 2020 Jun;45(2):75-86 [FREE Full text] [doi: 10.1007/s10484-020-09458-z] [Medline: 32246229]

XSL•FO
RenderX

50. Hung S, Li MS, Chen YL, Chiang JH, Chen YY, Hung GC. Smartphone-based ecological momentary assessment for Chinese patients with depression: an exploratory study in Taiwan. Asian J Psychiatr 2016 Oct;23:131-136. [doi: 10.1016/j.ajp.2016.08.003] [Medline: 27969071]
51. Deady M, Glozier N, Calvo R, Johnston D, Mackinnon A, Milne D, et al. Preventing depression using a smartphone app: a randomized controlled trial. Psychol Med 2022 Feb;52(3):457-466. [doi: 10.1017/S0033291720002081] [Medline: 32624013]
52. 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 Jan 11;5(1):e4 [FREE Full text] [doi: 10.2196/mental.9092] [Medline: 29326098]
53. Collins DA, Harvey SB, Lavender I, Glozier N, Christensen H, Deady M. A pilot evaluation of a smartphone application for workplace depression. Int J Environ Res Public Health 2020 Sep 16;17(18):6753 [FREE Full text] [doi: 10.3390/ijerph17186753] [Medline: 32947994]
54. Bonet L, Torous J, Arce D, Blanquer I, Sanjuan J. ReMindCare app for early psychosis: pragmatic real world intervention and usability study. JMIR Mhealth Uhealth 2020 Nov 06;8(11):e22997 [FREE Full text] [doi: 10.2196/22997] [Medline: 33155986]
55. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: a pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
56. Ryan KA, Babu P, Easter R, Saunders E, Lee AJ, Klasnja P, et al. A smartphone app to monitor mood symptoms in bipolar disorder: development and usability study. JMIR Ment Health 2020 Sep 22;7(9):e19476 [FREE Full text] [doi: 10.2196/19476] [Medline: 32960185]
57. Steare T, O'Hanlon P, Eskinazi M, Osborn D, Lloyd-Evans B, Jones R, et al. Smartphone-delivered self-management for first-episode psychosis: the ARIES feasibility randomised controlled trial. BMJ Open 2020 Aug 26;10(8):e034927 [FREE Full text] [doi: 10.1136/bmjopen-2019-034927] [Medline: 32847902]
58. Hidalgo-Mazzei D, Reinares M, Mateu A, Nikolova VL, Del Mar Bonnín C, 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] [Medline: 30145515]
59. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
60. Porras-Segovia A, Molina-Madueño RM, Berrouiguet S, López-Castroman J, Barrigón ML, Pérez-Rodríguez MS, et al. Smartphone-based ecological momentary assessment (EMA) in psychiatric patients and student controls: a real-world feasibility study. J Affect Disord 2020 Sep 01;274:733-741. [doi: 10.1016/j.jad.2020.05.067] [Medline: 32664009]
61. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
62. Beard C, Ramadurai R, McHugh RK, Pollak JP, Björgvinsson T. HabitWorks: development of a CBM-I smartphone app to augment and extend acute treatment. Behav Ther 2021 Mar;52(2):365-378. [doi: 10.1016/j.beth.2020.04.013] [Medline: 33622506]
63. Bell L, Garnett C, Qian T, Perski O, Williamson E, Potts HW. Engagement with a behavior change app for alcohol reduction: data visualization for longitudinal observational study. J Med Internet Res 2020 Dec 11;22(12):e23369 [FREE Full text] [doi: 10.2196/23369] [Medline: 33306026]
64. Laurens MC, Pieterse ME, Brusse-Keizer M, Salemink E, Ben Allouch S, Bohlmeijer ET, et al. Alcohol avoidance training as a mobile app for problem drinkers: longitudinal feasibility study. JMIR Mhealth Uhealth 2020 Apr 14;8(4):e16217 [FREE Full text] [doi: 10.2196/16217] [Medline: 32286235]
65. Herbec A, Brown J, Shahab L, West R, Raupach T. Pragmatic randomised trial of a smartphone app (NRT2Quit) to improve effectiveness of nicotine replacement therapy in a quit attempt by improving medication adherence: results of a prematurely terminated study. Trials 2019 Sep 02;20(1):547 [FREE Full text] [doi: 10.1186/s13063-019-3645-4] [Medline: 31477166]
66. Pallejà-Millán M, Rey-Reñones C, Barrera Uriarte ML, Granado-Font E, Basora J, Flores-Mateo G, et al. Evaluation of the Tobbstop mobile app for smoking cessation: cluster randomized controlled clinical trial. JMIR Mhealth Uhealth 2020 Jun 26;8(6):e15951 [FREE Full text] [doi: 10.2196/15951] [Medline: 32589153]
67. Zeng EY, Heffner JL, Copeland WK, Mull KE, Bricker JB. Get with the program: adherence to a smartphone app for smoking cessation. Addict Behav 2016 Dec;63:120-124 [FREE Full text] [doi: 10.1016/j.addbeh.2016.07.007] [Medline: 27454354]
68. Vilardaga R, Rizo J, Ries RK, Kientz JA, Ziedonis DM, Hernandez K, et al. Formative, multimethod case studies of learn to quit, an acceptance and commitment therapy smoking cessation app designed for people with serious mental illness. Transl Behav Med 2019 Nov 25;9(6):1076-1086 [FREE Full text] [doi: 10.1093/tbm/iby097] [Medline: 30445507]

XSL•FO
RenderX

69. Webb J, Peerbux S, Smittenaar P, Siddiqui S, Sherwani Y, Ahmed M, et al. Preliminary outcomes of a digital therapeutic intervention for smoking cessation in adult smokers: randomized controlled trial. JMIR Ment Health 2020 Oct 06;7(10):e22833 [FREE Full text] [doi: 10.2196/22833] [Medline: 33021488]
70. Hébert ET, Ra CK, Alexander AC, Helt A, Moisiuc R, Kendzor DE, et al. A mobile just-in-time adaptive intervention for smoking cessation: pilot randomized controlled trial. J Med Internet Res 2020 Mar 09;22(3):e16907 [FREE Full text] [doi: 10.2196/16907] [Medline: 32149716]
71. Witkiewitz K, Desai SA, Bowen S, Leigh BC, Kirouac M, Larimer ME. Development and evaluation of a mobile intervention for heavy drinking and smoking among college students. Psychol Addict Behav 2014 Sep;28(3):639-650 [FREE Full text] [doi: 10.1037/a0034747] [Medline: 25000269]
72. Chang AR, Bailey-Davis L, Hetherington V, Ziegler A, Yule C, Kwiecen S, et al. Remote dietary counseling using smartphone applications in patients with stages 1-3a chronic kidney disease: a mixed methods feasibility study. J Ren Nutr 2020 Jan;30(1):53-60 [FREE Full text] [doi: 10.1053/j.jrn.2019.03.080] [Medline: 31078403]
73. Chen YS, Wong JE, Ayob AF, Othman NE, Poh BK. Can Malaysian young adults report dietary intake using a food diary mobile application? A pilot study on acceptability and compliance. Nutrients 2017 Jan 13;9(1):62 [FREE Full text] [doi: 10.3390/nu9010062] [Medline: 28098770]
74. Hendrie GA, Hussain MS, Brindal E, James-Martin G, Williams G, Crook A. Impact of a mobile phone app to increase vegetable consumption and variety in adults: large-scale community cohort study. JMIR Mhealth Uhealth 2020 Apr 17;8(4):e14726 [FREE Full text] [doi: 10.2196/14726] [Medline: 32301739]
75. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014 Apr 15;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
76. Orlemann T, Reljic D, Zenker B, Meyer J, Eskofier B, Thiemt J, et al. A novel mobile phone app (OncoFood) to record and optimize the dietary behavior of oncologic patients: pilot study. JMIR Cancer 2018 Nov 20;4(2):e10703 [FREE Full text] [doi: 10.2196/10703] [Medline: 30459139]
77. Robinson E, Higgs S, Daley AJ, Jolly K, Lycett D, Lewis A, et al. Development and feasibility testing of a smart phone based attentive eating intervention. BMC Public Health 2013 Jul 09;13:639 [FREE Full text] [doi: 10.1186/1471-2458-13-639] [Medline: 23837771]
78. Bentley CL, Powell L, Potter S, Parker J, Mountain GA, Bartlett YK, et al. The use of a smartphone app and an activity tracker to promote physical activity in the management of chronic obstructive pulmonary disease: randomized controlled feasibility study. JMIR Mhealth Uhealth 2020 Jun 03;8(6):e16203 [FREE Full text] [doi: 10.2196/16203] [Medline: 32490838]
79. Edney S, Ryan JC, Olds T, Monroe C, Fraysse F, Vandelanotte C, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645 [FREE Full text] [doi: 10.2196/14645] [Medline: 31774402]
80. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
81. 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 Jun 12;54(7):518-528 [FREE Full text] [doi: 10.1093/abm/kaaa002] [Medline: 32182353]
82. Landers MR, Ellis TD. A mobile app specifically designed to facilitate exercise in Parkinson disease: single-cohort pilot study on feasibility, safety, and signal of efficacy. JMIR Mhealth Uhealth 2020 Oct 05;8(10):e18985 [FREE Full text] [doi: 10.2196/18985] [Medline: 33016887]
83. Luhanga ET, Hippocrate AA, Suwa H, Arakawa Y, Yasumoto K. Identifying and evaluating user requirements for smartphone group fitness applications. IEEE Access 2018 Jan 15;6:3256-3269. [doi: 10.1109/access.2018.2793844]
84. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
85. Bennett GG, Steinberg D, Askew S, Levine E, Foley P, Batch BC, et al. Effectiveness of an app and provider counseling for obesity treatment in primary care. Am J Prev Med 2018 Dec;55(6):777-786 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.005] [Medline: 30361140]
86. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
87. Forman EM, Goldstein SP, Crochiere RJ, Butryn ML, Juarascio AS, Zhang F, et al. Randomized controlled trial of OnTrack, a just-in-time adaptive intervention designed to enhance weight loss. Transl Behav Med 2019 Nov 25;9(6):989-1001. [doi: 10.1093/tbm/ibz137] [Medline: 31602471]

XSL•FO
RenderX

88. Hales S, Turner-McGrievy GM, Wilcox S, Fahim A, Davis RE, Huhns M, et al. Social networks for improving healthy weight loss behaviors for overweight and obese adults: a randomized clinical trial of the social pounds off digitally (Social POD) mobile app. Int J Med Inform 2016 Oct;94:81-90. [doi: 10.1016/j.ijmedinf.2016.07.003] [Medline: 27573315]
89. Ifejika NL, Bhadane M, Cai CC, Noser EA, Grotta JC, Savitz SI. Use of a smartphone-based mobile app for weight management in obese minority stroke survivors: pilot randomized controlled trial with open blinded end point. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e17816 [FREE Full text] [doi: 10.2196/17816] [Medline: 32319963]
90. 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 [FREE Full text] [doi: 10.1097/JCN.0000000000000307] [Medline: 26646593]
91. Svetkey LP, Batch BC, Lin PH, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141 [FREE Full text] [doi: 10.1002/oby.21226] [Medline: 26530929]
92. Tanaka K, Sasai H, Wakaba K, Murakami S, Ueda M, Yamagata F, et al. Professional dietary coaching within a group chat using a smartphone application for weight loss: a randomized controlled trial. J Multidiscip Healthc 2018 Jul 16;11:339-347 [FREE Full text] [doi: 10.2147/JMDH.S165422] [Medline: 30038502]
93. Horsch C, Spruit S, Lancee J, van Eijk R, Beun RJ, Neerincx M, et al. Reminders make people adhere better to a self-help sleep intervention. Health Technol (Berl) 2017;7(2):173-188 [FREE Full text] [doi: 10.1007/s12553-016-0167-x] [Medline: 29201588]
94. 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] [Medline: 28400355]
95. Göransson C, Wengström Y, Hälleberg-Nyman M, Langius-Eklöf A, Ziegert K, Blomberg K. An app for supporting older people receiving home care - usage, aspects of health and health literacy: a quasi-experimental study. BMC Med Inform Decis Mak 2020 Sep 15;20(1):226 [FREE Full text] [doi: 10.1186/s12911-020-01246-3] [Medline: 32933500]
96. Murawski B, Plotnikoff RC, Rayward AT, Oldmeadow C, Vandelanotte C, Brown WJ, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
97. Laranjo L, Quiroz JC, Tong HL, Arevalo Bazalar M, Coiera E. A mobile social networking app for weight management and physical activity promotion: results from an experimental mixed methods study. J Med Internet Res 2020 Dec 08;22(12):e19991 [FREE Full text] [doi: 10.2196/19991] [Medline: 33289670]
98. Oftedal S, Burrows T, Fenton S, Murawski B, Rayward AB, Duncan MJ. Feasibility and preliminary efficacy of an m-health intervention targeting physical activity, diet, and sleep quality in shift-workers. Int J Environ Res Public Health 2019 Oct 10;16(20):3810 [FREE Full text] [doi: 10.3390/ijerph16203810] [Medline: 31658624]
99. Recio-Rodriguez JI, Agudo-Conde C, Martin-Cantera C, González-Viejo MN, Fernandez-Alonso MD, Arietaleanizbeaskoa MS, EVIDENT Investigators. Short-term effectiveness of a mobile phone app for increasing physical activity and adherence to the Mediterranean diet in primary care: a randomized controlled trial (EVIDENT II Study). J Med Internet Res 2016 Dec 19;18(12):e331 [FREE Full text] [doi: 10.2196/jmir.6814] [Medline: 27993759]
100. Seng EK, Prieto P, Boucher G, Vives-Mestres M. Anxiety, incentives, and adherence to self-monitoring on a mobile health platform: a naturalistic longitudinal cohort study in people with headache. Headache 2018 Nov;58(10):1541-1555 [FREE Full text] [doi: 10.1111/head.13422] [Medline: 30334248]
101. Avalos LA, Aghaee S, Kurtovich E, Quesenberry Jr C, Nkemere L, McGinnis MK, et al. A mobile health mindfulness intervention for women with moderate to moderately severe postpartum depressive symptoms: feasibility study. JMIR Ment Health 2020 Nov 12;7(11):e17405 [FREE Full text] [doi: 10.2196/17405] [Medline: 33180028]
102. 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 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
103. Chandler J, Sox L, Diaz V, Kellam K, Neely A, Nemeth L, et al. Impact of 12-month smartphone breathing meditation program upon systolic blood pressure among non-medicated stage 1 hypertensive adults. Int J Environ Res Public Health 2020 Mar 17;17(6):1955 [FREE Full text] [doi: 10.3390/ijerph17061955] [Medline: 32192020]
104. 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 Dec 31;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
105. Goldberg SB, Imhoff-Smith T, Bolt DM, Wilson-Mendenhall CD, Dahl CJ, Davidson RJ, et al. Testing the efficacy of a multicomponent, self-guided, smartphone-based meditation app: three-armed randomized controlled trial. JMIR Ment Health 2020 Nov 27;7(11):e23825 [FREE Full text] [doi: 10.2196/23825] [Medline: 33245288]
106. Huberty J, Vranceanu AM, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]

XSL•FO
RenderX

107. Lengacher CA, Reich RR, Ramesar S, Alinat CB, Moscoso M, Cousin L, et al. Feasibility of the mobile mindfulness-based stress reduction for breast cancer (mMBSR(BC)) program for symptom improvement among breast cancer survivors. Psychooncology 2018 Feb;27(2):524-531 [FREE Full text] [doi: 10.1002/pon.4491] [Medline: 28665541]
108. Mahlo L, Windsor TD. Feasibility, acceptability, and preliminary efficacy of an app-based mindfulness-meditation program among older adults. Gerontologist 2021 Jul 13;61(5):775-786. [doi: 10.1093/geront/gnaa093] [Medline: 32663286]
109. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 08;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]
110. Aji M, Glozier N, Bartlett D, Peters D, Calvo RA, Zheng Y, et al. A feasibility study of a mobile app to treat insomnia. Transl Behav Med 2021 Mar 16;11(2):604-612. [doi: 10.1093/tbm/ibaa019] [Medline: 32227087]
111. Gentili C, Zetterqvist V, Rickardsson J, Holmström L, Simons LE, Wicksell RK. ACTsmart: guided smartphone-delivered acceptance and commitment therapy for chronic pain-a pilot trial. Pain Med 2021 Feb 23;22(2):315-328 [FREE Full text] [doi: 10.1093/pm/pnaa360] [Medline: 33200214]
112. 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 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
113. Kang SG, Kang JM, Cho SJ, Ko KP, Lee YJ, Lee HJ, et al. Cognitive behavioral therapy using a mobile application synchronizable with wearable devices for insomnia treatment: a pilot study. J Clin Sleep Med 2017 Apr 15;13(4):633-640 [FREE Full text] [doi: 10.5664/jcsm.6564] [Medline: 28162145]
114. Kelechi TJ, Prentice MA, Mueller M, Madisetti M, Vertegel A. A lower leg physical activity intervention for individuals with chronic venous leg ulcers: randomized controlled trial. JMIR Mhealth Uhealth 2020 May 15;8(5):e15015 [FREE Full text] [doi: 10.2196/15015] [Medline: 32412419]
115. 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]
116. Minen MT, Schaubhut KB, Morio K. Smartphone based behavioral therapy for pain in multiple sclerosis (MS) patients: a feasibility acceptability randomized controlled study for the treatment of comorbid migraine and ms pain. Mult Scler Relat Disord 2020 Nov;46:102489 [FREE Full text] [doi: 10.1016/j.msard.2020.102489] [Medline: 32950893]
117. Ong SW, Jassal SV, Miller JA, Porter EC, Cafazzo JA, Seto E, et al. Integrating a smartphone-based self-management system into usual care of advanced CKD. Clin J Am Soc Nephrol 2016 Jun 06;11(6):1054-1062 [FREE Full text] [doi: 10.2215/CJN.10681015] [Medline: 27173169]
118. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2020;32(2):109-116. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
119. Weerts ZZ, Heinen KG, Masclee AA, Quanjel AB, Winkens B, Vork L, et al. Smart data collection for the assessment of treatment effects in irritable bowel syndrome: observational study. JMIR Mhealth Uhealth 2020 Nov 02;8(11):e19696 [FREE Full text] [doi: 10.2196/19696] [Medline: 33030150]
120. Isetta V, Torres M, González K, Ruiz C, Dalmases M, Embid C, et al. A new mHealth application to support treatment of sleep apnoea patients. J Telemed Telecare 2017 Jan;23(1):14-18. [doi: 10.1177/1357633X15621848] [Medline: 26672606]
121. Zia J, Schroeder J, Munson S, Fogarty J, Nguyen L, Barney P, et al. Feasibility and usability pilot study of a novel irritable bowel syndrome food and gastrointestinal symptom journal smartphone app. Clin Transl Gastroenterol 2016 Mar 03;7(3):e147 [FREE Full text] [doi: 10.1038/ctg.2016.9] [Medline: 26938478]
122. Choi JY, Choi H, Seomun G, Kim EJ. Mobile-application-based interventions for patients with hypertension and ischemic heart disease: a systematic review. J Nurs Res 2020 Oct;28(5):e117. [doi: 10.1097/JNR.0000000000000382] [Medline: 32501962]
123. Linardon J, Shatte A, Rosato J, Fuller-Tyszkiewicz M. Efficacy of a transdiagnostic cognitive-behavioral intervention for eating disorder psychopathology delivered through a smartphone app: a randomized controlled trial. Psychol Med (forthcoming) 2020 Sep 25:1-12. [doi: 10.1017/S0033291720003426] [Medline: 32972467]
124. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
125. Wang C, Qi H. Influencing factors of acceptance and use behavior of mobile health application users: systematic review. Healthcare (Basel) 2021 Mar 22;9(3):357 [FREE Full text] [doi: 10.3390/healthcare9030357] [Medline: 33809828]
126. Mishra V, Künzler F, Kramer JN, Fleisch E, Kowatsch T, Kotz D. Detecting receptivity for mHealth interventions in the natural environment. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Jun;5(2):74 [FREE Full text] [doi: 10.1145/3463492] [Medline: 34926979]
127. Künzler F, Mishra V, Kramer JN, Kotz D, Fleisch E, Kowatsch T. Exploring the state-of-receptivity for mHealth interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Dec 11;3(4):1-27 [FREE Full text] [doi: 10.1145/3369805]

XSL•FO
RenderX

128. Fortuna KL, Muralidharan A, Goldstein CM, Venegas M, Glass JE, Brooks JM. Certified peer specialists' perspective of the barriers and facilitators to mobile health engagement. J Technol Behav Sci 2020 Dec;5(4):318-323 [FREE Full text] [doi: 10.1007/s41347-020-00138-7] [Medline: 33163620]
129. Rubin A, Livingston NA, Brady J, Hocking E, Bickmore T, Sawdy M, et al. Computerized relational agent to deliver alcohol brief intervention and referral to treatment in primary care: a randomized clinical trial. J Gen Intern Med 2022 Jan;37(1):70-77 [FREE Full text] [doi: 10.1007/s11606-021-06945-9] [Medline: 34145518]
130. Kowatsch T, Lohse KM, Erb V, Schittenhelm L, Galliker H, Lehner R, et al. Hybrid ubiquitous coaching with a novel combination of mobile and holographic conversational agents targeting adherence to home exercises: four design and evaluation studies. J Med Internet Res 2021 Feb 22;23(2):e23612 [FREE Full text] [doi: 10.2196/23612] [Medline: 33461957]
131. 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 Feb 17;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
132. Abdullah AS, Gaehde S, Bickmore T. A tablet based embodied conversational agent to promote smoking cessation among veterans: a feasibility study. J Epidemiol Glob Health 2018 Dec;8(3-4):225-230 [FREE Full text] [doi: 10.2991/j.jegh.2018.08.104] [Medline: 30864768]
133. De Croon R, Geuens J, Verbert K, Vanden AV. A systematic review of the effect of gamification on adherence across disciplines. In: Proceedings of the 3rd International Conference on HCI in Games: Experience Design and Game Mechanics. Washington, DC, USA: Springer International Publishing; 2021 Jul Presented at: HCI-Games '21; July 24-29, 2021; Virtual p. 168-184. [doi: 10.1007/978-3-030-77277-2_14]
134. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: a randomized controlled trial. PLoS One 2020 Sep 2;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
135. Blok AC, Sadasivam RS, Amante DJ, Kamberi A, Flahive J, Morley J, et al. Gamification to motivate the unmotivated smoker: the "take a break" digital health intervention. Games Health J 2019 Aug;8(4):275-284 [FREE Full text] [doi: 10.1089/g4h.2018.0076] [Medline: 31219347]
136. Boyle SC, Earle AM, LaBrie JW, Smith DJ. PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, web-based personalized normative feedback alcohol interventions. Addict Behav 2017 Apr;67:8-17 [FREE Full text] [doi: 10.1016/j.addbeh.2016.11.024] [Medline: 27978426]
137. Harris MA. Maintenance of behaviour change following a community-wide gamification based physical activity intervention. Prev Med Rep 2018 Nov 3;13:37-40 [FREE Full text] [doi: 10.1016/j.pmedr.2018.11.009] [Medline: 30510892]
138. González-González C, Río NG, Navarro-Adelantado V. Exploring the benefits of using gamification and videogames for physical exercise: a review of state of art. Int J Interact Multimed Artif Intell 2018 Sep;5(2):46-52. [doi: 10.9781/ijimai.2018.03.005]
139. Abdul Rahim MI, Thomas RH. Gamification of medication adherence in epilepsy. Seizure 2017 Nov;52:11-14 [FREE Full text] [doi: 10.1016/j.seizure.2017.09.008] [Medline: 28934624]
140. Johnson D, Deterding S, Kuhn KA, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
141. Floryan M, Chow PI, Schueller SM, Ritterband LM. The model of gamification principles for digital health interventions: evaluation of validity and potential utility. J Med Internet Res 2020 Jun 10;22(6):e16506 [FREE Full text] [doi: 10.2196/16506] [Medline: 32519965]
142. Sampat BH, Prabhakar B. Privacy risks and security threats in mHealth apps. J Int Technol Inf Manag 2017;26(4):126-153 [FREE Full text]
143. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst 2015 Jan;39(1):181. [doi: 10.1007/s10916-014-0181-3] [Medline: 25486895]
144. Hutton L, Price BA, Kelly R, McCormick C, Bandara AK, Hatzakis T, et al. Assessing the privacy of mHealth apps for self-tracking: heuristic evaluation approach. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e185 [FREE Full text] [doi: 10.2196/mhealth.9217] [Medline: 30348623]
145. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8 [FREE Full text] [doi: 10.2196/mhealth.3672] [Medline: 25599627]
146. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
147. 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] [Medline: 32560936]

XSL•FO
RenderX

148. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research . . . and three things we can do about them. Psychiatr Serv 2017 May 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
149. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
150. Kohl LF, Crutzen R, de Vries NK. Online prevention aimed at lifestyle behaviors: a systematic review of reviews. J Med Internet Res 2013 Jul 16;15(7):e146 [FREE Full text] [doi: 10.2196/jmir.2665] [Medline: 23859884]

Abbreviations

CVD: cardiovascular disease
DHI: digital health intervention
mHealth: mobile health
NCD: noncommunicable disease
PICOS: Population, Intervention, Comparison, Outcomes, and Study component
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UX: user experience

Edited by A Mavragani; submitted 02.12.21; peer-reviewed by M Wu, F Velayati, H Meng; comments to author 23.12.21; revised version received 31.03.22; accepted 09.04.22; published 25.05.22.

Please cite as:
Jakob R, Harperink S, Rudolf AM, Fleisch E, Haug S, Mair JL, Salamanca-Sanabria A, Kowatsch T
Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review
J Med Internet Res 2022;24(5):e35371
URL: https://www.jmir.org/2022/5/e35371
doi:10.2196/35371
PMID:35612886

©Robert Jakob, Samira Harperink, Aaron Maria Rudolf, Elgar Fleisch, Severin Haug, Jacqueline Louise Mair, Alicia Salamanca-Sanabria, Tobias Kowatsch. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2022. 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.

XSL•FO
RenderX

Review

Findings and Guidelines on Provider Technology, Fatigue, and Well-being: Scoping Review

Donald M Hilty[1,2], MBA, MD; Christina M Armstrong[3], PhD; Shelby A Smout[4], MS; Allison Crawford[5], MD, PhD; Marlene M Maheu[6], PhD; Kenneth P Drude[7], PhD; Steven Chan[8], MBA, MD; Peter M Yellowlees[1], MBBS, MD; Elizabeth A Krupinski[9], PhD

[1]Department of Psychiatry & Behavioral Sciences, University of California Davis School of Medicine, Sacramento, CA, United States

[2]Northern California Veterans Affairs Health Care System, Mather, CA, United States

[3]Office of Connected Care, Department of Veterans Affairs, Washington, DC, United States

[4]Virginia Commonwealth University, Richmond, VA, United States

[5]Extension for Community Healthcare Outcomes, Ontario Mental Health at Centre for Addiction and Mental Health, University of Toronto Virtual Mental Health, and Canada Suicide Prevention Service, Toronto, ON, Canada

[6]Telebehavioral Health Institute, LLC and Coalition for Technology in Behavioral Science, San Diego, CA, United States

[7]Coalition for Technology in Behavioral Science & Ohio Board of Psychology, Dayton, OH, United States

[8]Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine & Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, United States

[9]Department of Radiology & Imaging Sciences, Emory University School of Medicine, Atlanta, GA, United States

Corresponding Author:
Donald M Hilty, MBA, MD
Department of Psychiatry & Behavioral Sciences
University of California Davis School of Medicine
2230 Stockton Boulevard
Sacramento, CA, 95817
United States
Phone: 1 626 375 7857
Email: donh032612@gmail.com

Abstract

Background: Video and other technologies are reshaping the delivery of health care, yet barriers related to workflow and possible provider fatigue suggest that a thorough evaluation is needed for quality and process improvement.

Objective: This scoping review explored the relationship among technology, fatigue, and health care to improve the conditions for providers.

Methods: A 6-stage scoping review of literature (from 10 databases) published from 2000 to 2020 that focused on technology, health care, and fatigue was conducted. Technologies included synchronous video, telephone, informatics systems, asynchronous wearable sensors, and mobile health devices for health care in 4 concept areas related to provider experience: behavioral, cognitive, emotional, and physical impact; workplace at the individual, clinic, hospital, and system or organizational levels; well-being, burnout, and stress; and perceptions regarding technology. Qualitative content, discourse, and framework analyses were used to thematically analyze data for developing a spectrum of health to risk of fatigue to manifestations of burnout.

Results: Of the 4221 potential literature references, 202 (4.79%) were duplicates, and our review of the titles and abstracts of 4019 (95.21%) found that 3837 (90.9%) were irrelevant. A full-text review of 182 studies revealed that 12 (6.6%) studies met all the criteria related to technology, health care, and fatigue, and these studied the behavioral, emotional, cognitive, and physical impact of workflow at the individual, hospital, and system or organizational levels. Video and electronic health record use has been associated with physical eye fatigue; neck pain; stress; tiredness; and behavioral impacts related to additional effort owing to barriers, trouble with engagement, emotional wear and tear and exhaustion, cognitive inattention, effort, expecting problems, multitasking and workload, and emotional experiences (eg, anger, irritability, stress, and concern about well-being). An additional 14 studies that evaluated behavioral, emotional, and cognitive impacts without focusing on fatigue found high user ratings on data quality, accuracy, and processing but low satisfaction with clerical tasks, the effort required in work, and interruptions costing time, resulting in more errors, stress, and frustration. Our qualitative analysis suggests a spectrum from health to risk and provides

XSL•FO
RenderX

an outline of organizational approaches to human factors and technology in health care. Business, occupational health, human factors, and well-being literature have not studied technology fatigue and burnout; however, their findings help contextualize technology-based fatigue to suggest guidelines. Few studies were found to contextually evaluate differences according to health professions and practice contexts.

Conclusions: Health care systems need to evaluate the impact of technology in accordance with the Quadruple Aim to support providers' well-being and prevent workload burden, fatigue, and burnout. Implementation and effectiveness approaches and a multilevel approach with objective measures for clinical, human factors, training, professional development, and administrative workflow are suggested. This requires institutional strategies and competencies to integrate health care quality, technology and well-being outcomes.

(J Med Internet Res 2022;24(5):e34451) doi:10.2196/34451

KEYWORDS

burnout; screen fatigue; technology fatigue; well-being; videoconferencing; Zoom fatigue; mobile phone

Introduction

Background

Technology is reshaping the delivery of health care worldwide as a facilitator, practice extender, and virtual team member for person- and patient-centered care [1,2]. Health care systems and governmental agencies worldwide are promoting quality and evidence-based care by implementing the Quadruple Aim, which emphasizes reducing costs and improving population health, patient experience, and team well-being [1-3]. Fatigue is a very complex and multidimensional construct and a review of research across cognitive science, exercise physiology, and clinical practice suggests that its most promising common feature is the notion of *perceived effort*—this accounts for interindividual differences and situational variations and includes both mental and physical constructs and integrates motivational and emotional dimensions [4]. Health care providers and other employees have increasingly noted problems related to fatigue and excessive workflow steps, particularly electronic health records (EHRs), that may affect well-being and contribute to burnout [5].

Technology challenges include learning to use it personally and professionally and integrating it into workflow and screen time [6-8]. Subjective phrases such as *technology fatigue* or, in the COVID-19 era, *Zoom fatigue* suggest that technology *causes* fatigue. Past research on employees' subjective, physical, and ophthalmologic factors related to computer displays [9-14] suggests that there are many additional occupational health factors related to fatigue, burnout, and accidents [15-17]. Business industries have contended with technological challenges and systematically changed workflows for users to transform and avoid extinction [18]. In health care, there appears to be a gap between the system's perception of processes and users' or participants' experiences [2].

Current Practice

It appears that there is a gap in how health care providers typically use EHRs and other technologies, the amount of effort required for workflow, and how this leads to fatigue (or burnout). Health care is starting to evaluate the longitudinal continua of work engagement and burnout, the development of burnout in relation to job demands and resources, and the role of psychosocial working conditions [19,20]. Assessment of well-being [21,22], burnout [23-28], burnout with EHRs [29,30], and related risk factors [31] is underway, including in psychiatry and behavioral health providers [32-34]. Interventions can help prevent and ameliorate burnout [35,36] and changes to organizational structure (eg, shared leadership), process improvement (eg, lean), employee support (eg, leisure, fitness, and diet), and professional development [37,38]. Another gap is that systems have generally approached burnout as an individual's problem (eg, depression) rather than as an organization's problem (ie, a shared problem). Key stressors within an organization that put people at risk of burnout need to be identified—at a department or unit level—so that changes can be made to reduce their impact and create healthier workplaces.

Objective

The relationship among technology, fatigue, and health care can be better understood by reviewing the broad literature on health, business, occupational health, technology, and well-being. This will help with the following:

1. Find data on the intersection of technology, fatigue, and health care (eg, association, mediation, and cause).
2. Provide an overview of the business, occupational health, and well-being literature to contextualize technology-based fatigue, its components, and related processes.
3. Suggest guidelines for health care related to technology, well-being, and fatigue at provider, clinic, and system levels to advance self-assessment, quality improvement, and necessary organizational and social improvements to promote a culture of well-being.

Methods

Approach

A literature search via the Medical Subject Headings of the keywords spanned from January 2000 to December 2020, using the original 6-stage scoping review process [39], with updated modifications [40] and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) [41].

Research Question

This scoping review explores the relationship among technology, fatigue, and health care to improve the conditions for providers.

XSL•FO
RenderX

It focuses on the overarching question: "What is technology-based fatigue and what are its consequences for providers and patients?" The subquestions are as follows:

1. What are the characteristics of technology-based fatigue and its associated factors, including technologies?
2. Does technology and associated fatigue impact provider health (burnout, compassion fatigue, and well-being)?
3. How does provider burnout or well-being associated with technology affect the delivery of care; therapeutic relationships; and quality of care offered in person, by video, and by other technologies?
4. What are strategies or interventions being used to prevent or ameliorate technology fatigue?

The goal was to synthesize clinical, provider, administrative, business, and other workplace data and consider the current and target states for using technologies in a healthy way to prevent or minimize problems and focus efforts on further assessment and intervention.

Identifying Relevant Studies

A total of 10 literature databases were queried by a librarian: PubMed, APA PsycNET, Embase, PsycINFO database via the Ebsco platform, Web of Science, Scopus, Social Sciences Citation Index, Telemedicine Information Exchange database, Centre for Reviews and Dissemination, and Cochrane Database of Systematic Reviews and Central Register of Controlled Trials. The search focused on technology, health care, and fatigue via synchronous telepsychiatry and tele-behavioral or tele-mental health, though telephone, asynchronous, mobile health, tablets, and text were also searched. It also included types of health providers (ie, clinician, provider, counselor, employee, medical nurse or physician, psychiatrist, psychologist, social worker, therapist, and worker), assessment, care, evaluation, screening, therapy, triage, and treatment. The initial search targeted 4 concept areas by using specific terms as shown in Textbox 1.

Textbox 1. Concept areas used in the initial search.

Behavioral, cognitive, emotional, and physical impact

- Behavioral impact
  - *(anxi*, barriers, boredom, complain*, concern*, depression *, detachment *, distance, effort*, engage*, emotional*, enjoy*, exhaustion, experience, factor, fatigue*, insomnia, intimacy, isolation, mental, onerous, positive, readiness, reward*, social, substance, suicide, team, worry)* or
- Cognitive impact
  - *(attention, attitude*, alertness, critical, cynicism, distraction, efficacy, effort, expectation, incompetence, indecision, motivation, multitasking, negative, step*, task*, workflow, workload)* or
- Emotional impact
  - *(alone, anger, anxiety, compassion*, complex, confidence, empower, esteem, human, irritability, lonely, positive, quality of life, resilien*, sadness, satisfaction, secondary, share*, trauma, satisfaction, stress, support, susceptible, therapeutic, wellness, well-being)* or
- Physical impact
  - *(ache, back, distress, exhaustion, eye, fatigue, headache, neck, pain, problem*, strain, stress, tiredness, visual)*

Workplace at the individual, clinic, hospital, and system or organizational levels

- *accessories, alternative, burden, clerical, computer, control, dedicated, demand, display, distraction, disrupti*, error*, flexib*, home, interruptions, intrusion, job, mishap, mistake, nap, organization, recognition, routine, relative value unit (RVU), safety, schedule, screen, separation, shift, telework, terminal, time, video, voice, workflow, and workload*

Well-being, burnout, and stress

- *adaptable, adjustment, burnout, confidence, cop*, esteem, fitness, happy, health*, mindful*, purposeful, relaxation, resilien*, risk, safety, satisfaction, vitality, vulnerab*, wellness, willingness*

Provider perceptions regarding technology

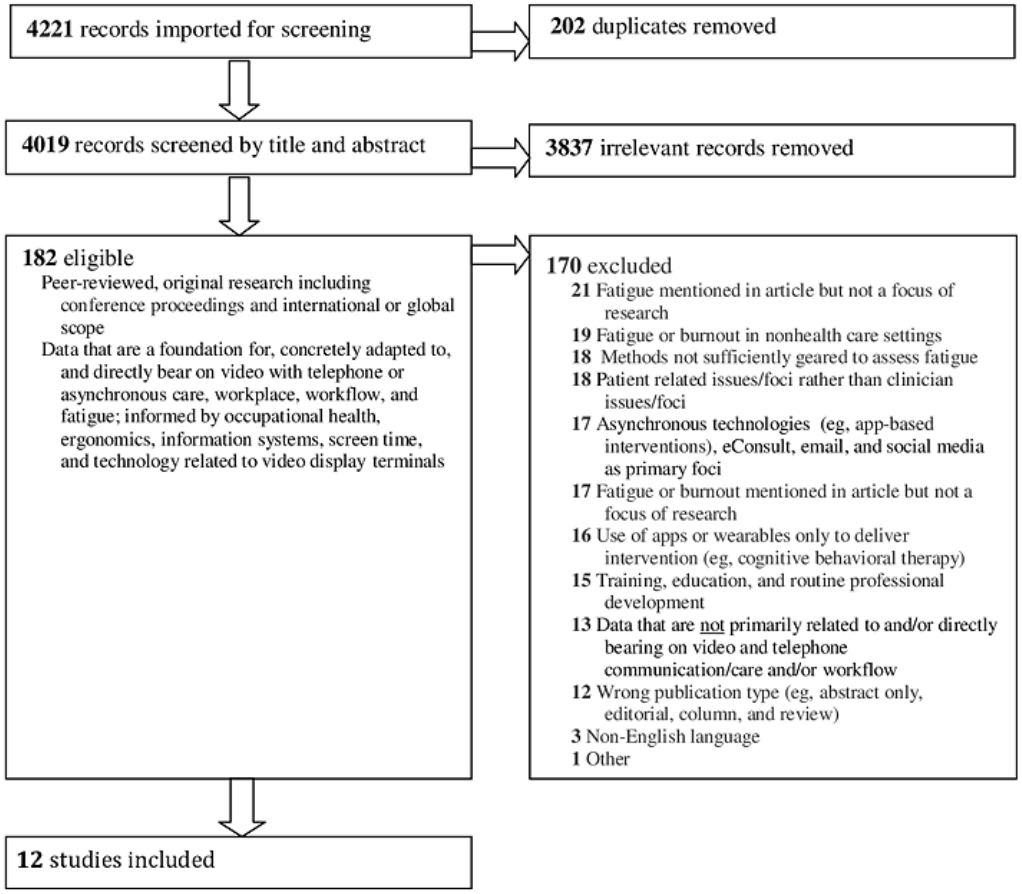
- *attitudes, diffusion, adaptor, and willingness, motivation, urgency, readiness to use technology, biases regarding tech use, and experience of using technology*

Study Selection

One author (DMH) screened titles and abstracts of 4221 potential references, excluding 202 (4.79%) based on duplication and 3837 (90.9%) that did not meet the search criteria. Notably, 2 of 3 authors (DMH, CMA, or SAS) reviewed the full text of 182 articles, but only 12 (6.6%) met the inclusion criteria related to health care, technology, and fatigue based on consensus (Figure 1). If there was a disagreement, a third author (DMH, CMA, or SAS) made the decision. An additional 14 studies evaluated health care and technology workflow with user experiences and perceptions that may provide a contextual understanding of fatigue.

XSL•FO
RenderX

Figure 1. Search flowchart: diagram of studies reviewed. eConsult: electronic consultation.



Data Charting

A data charting form was used to extract data, and notes were organized consistent with a descriptive analytical method by each reviewer. The reviewers compared and consolidated the information using a qualitative content analysis approach [42].

Analysis, Reporting, and the Meaning of Findings

Results were organized based on the objectives into tables and figures, with key concepts and components of technology-based fatigue outlined and described, partially based on excerpts from published topics. As this research area, although critical, is nascent, findings were reported individually.

The technologies used have evolved considerably, making these articles a challenge to compare. Qualitative steps to analyze disparate populations, data and methods of studies were used [42]. Content, discourse, and framework qualitative analysis techniques were used to analyze findings from papers to develop a spectrum of health to risk of fatigue to manifestations of burnout (Figure 2) [43]. Content analysis was used to classify, summarize, and tabulate the behavioral data; discourse analysis was used to search for themes and patterns; and framework analysis was used to sift through, chart, and sort data in accordance with key issues and themes in a series of steps (eg, indexing, charting, mapping, and interpretation). Time points related to the release and integration of new technologies into the marketplace and health care, as well as concept area terminology surfacing in the literature were estimated qualitatively (Figure 3).

XSL•FO
RenderX

Figure 2. A comparison of health and resilience, risk to well-being, and manifestations of technology-based fatigue and burnout. EHR: electronic health record.

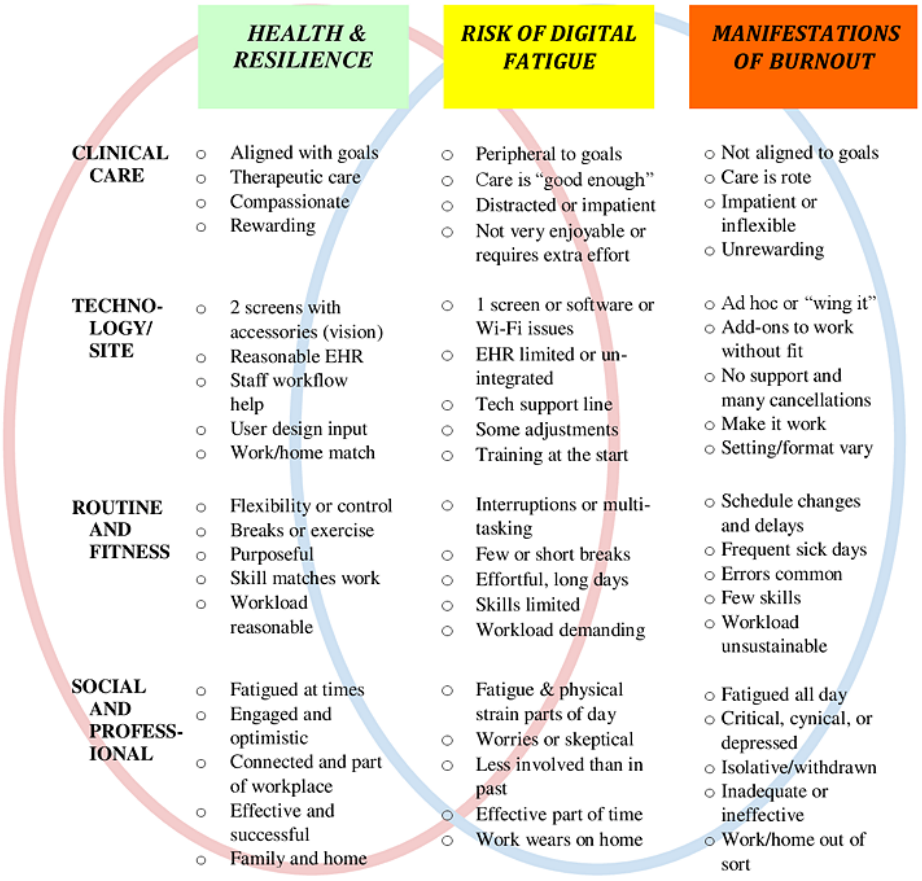
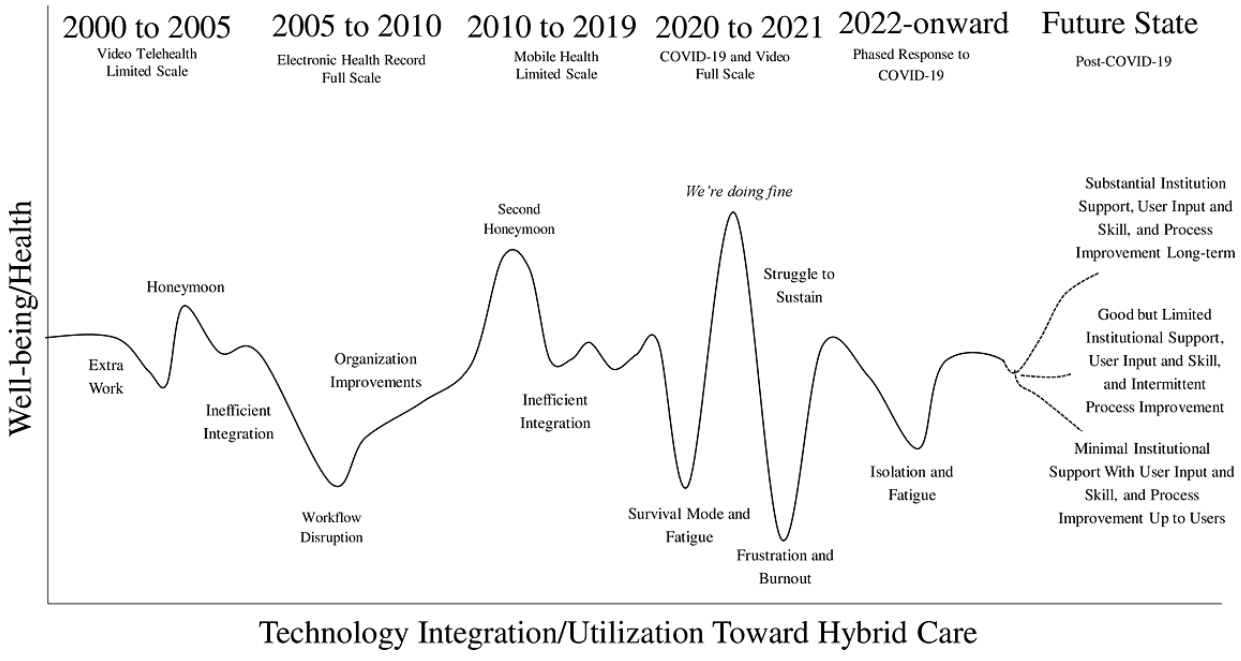


Figure 3. Impact of the implementation of technology integration and utilization toward hybrid care on health and well-being over time.



XSL•FO
RenderX

Expert Opinions

Expert opinions were solicited to review preliminary findings and suggest additional steps for improvement. A list of relevant experts was compiled from (1) behavioral health organizations internationally (eg, Psychiatry, Psychology, Social work, and Addiction); (2) technology-related special interest groups of organizations (eg, the American Telemedicine Association American Medical or Nursing Informatics Associations and Coalition for Technology in Behavioral Sciences); (3) health organizations related to quality improvement, human resources, occupational health, and lean systems (eg, Agency for Healthcare Research and Quality, American National Standards Institute, Healthcare Information and Management Systems Society, Joint Commission, and World Health Organization); and (4) federal (ie, US National Academy of Sciences, US. National Institute of Health, US Department of Defense, and US Veterans Health Administration) and academic institutions (ie, Mayo Clinic and University of California); and (5) researchers, authors, editors, and editorial board members.

Experts were invited by email from 7 countries (Australia, Canada, Germany, India, Italy, the United Kingdom, and the United States) by several means, including attending a live videoconference expert feedback session and providing qualitative feedback. The lead author (DMH) facilitated the use of a scribe, and each of the 3 sessions lasted 50 minutes. The abstract, objectives, methods, tables, and figures were sent to experts a week in advance. The session started with a brief introduction based on the abstract, objectives, and overview of the table and figure content (10 minutes). This was followed by general questions, comments, and suggestions, including review of the data charting and search criteria (20 minutes). The input was summarized and themes were extracted to guide the organization (eg, headings in rows) and content (eg, in the columns) of tables and figures. The questions were asked to solicit additional feedback (10 minutes), and other suggestions were provided at the end of the session. Feedback was collated based on previous studies using consensus and modified Delphi processes [37,44]. Attendees were asked to complete a qualitative and quantitative 5-item Likert-scale survey (strongly disagree, disagree, neutral, agree, or strongly agree) or provide qualitative feedback via email. The survey included 6 questions, 3 weighted positively (the data provide a systematic way for clinicians to assess fatigue and well-being) and 3 weighted negatively.

Results

Overview

The results are organized per objectives (intersection of technology, fatigue, and health care; business, occupational health, and well-being literature; and guidelines for health care), which align with the search of the 4 concept areas (behavioral, cognitive, emotional, and physical impact; workplace at the individual, clinic, hospital, and system or organizational levels; well-being, burnout, and stress; and provider perceptions regarding technology). Business, occupational health, and well-being literature did not study technology fatigue and burnout; however, findings help contextualize technology-based fatigue to suggest guidelines at provider, clinic, and system levels for health care. Few studies were found to contextually evaluate differences according to health professions and practice contexts.

Expert opinions and feedback (N=19) contributed by attending a live expert feedback session and providing qualitative feedback, completing a qualitative and quantitative 6-item Likert-scale survey (16/19, 84%), or providing qualitative feedback via email (2/19, 10%). Of the 19 attendees in expert feedback sessions, 8 (42%) were psychiatrists, 5 (26%) were psychologists, 2 (10%) were marriage and family therapists, 1 (5%) was a physician (not psychiatrist), 1 (5%) was a counselor, 1 (5%) was a social worker, and 1 (5%) was a systems engineer. Results showed that most attendees agreed or strongly agreed that (1) “The results provided in tables are organized in the ballpark and relatively complete” (18/19 93%), (2) “The tables are a practical way to identify, analyze, and begin to address technology problems for providers and systems” (13/19, 69%); and (3) “The figures substantially help to compare and contrast the continuum of health versus fatigue versus burnout” (13/19, 69%).

Technology, Health Care, and Fatigue

A total of 12 papers met the inclusion criteria based on the consensus of the authors [8,13,14,45-53]. Studies assessed the behavioral (8/12, 67%), emotional (4/12, 33%), cognitive (7/12, 58%), and physical (4/12, 33%) impact of workflow at the individual (11/12, 92%), clinic (8/12, 67%), hospital (6/12, 50%), and system or organizational (6/12, 50%) levels; only 25% (3/12) of studies included all levels. Most health care professionals had medical degrees (MD; 8/12, 67%), including radiologists (2/12, 17%). Video and EHR use was associated with behavioral, cognitive, emotional, and physical impact, with the latter usually reported as eye fatigue, neck pain, stress, and tiredness. Behavioral impact involved additional efforts regarding barriers, trouble with engagement, emotional wear and tear, exhaustion, and fatigue. Cognitive impact focused on inattention, effort, expecting problems, multitasking, and workload. Emotional impact was related to anger, irritability, stress, and concern about well-being.

These studies were conducted in the United States, although a study compared providers’ impact across countries. Only 17% (2/12) studies discussed the physical environment, occupational health approaches, mobile care, telework, or lean, human factors, and user design approaches to workflow. System onboarding and training enables users to get oriented and informally sets expectations, but often there are no processes for ongoing self-, peer-, and system-assessment of experience or skills. Workplace, workspace, ergonomic, and technology implementation are gaining more attention in health care [54,55] and other industries for those who function at work and home [56]. The studies were unidirectional in association, mediation, and causation—technology causing fatigue, and similar to other studies in the literature [11,12,57,58], they lacked standard assessment, monitoring, and interventions.

Studies have focused on the use [8,45,53], surveys of providers [46,51], visual strain or fatigue [13,14], implementation and usability [47,52], and consensus reports [59,60] (Multimedia

XSL•FO
RenderX

Appendix 1 [8,13,14,45-53]). It is organized by study, sample size (N), length of time, population, country, design, type of technology, area of focus of the assessment (ie, behavioral, cognitive, emotional, and physical impact), and level of the assessment (ie, individual, clinic, hospital, and system or organizational). Physician participants experienced physiological fatigue at least once during simulation exercises involving 4 patient cases, with the majority (20/25, 80%) experiencing physiological fatigue within the first 22 minutes of use [8]. Those who experienced EHR-related fatigue in a patient case were less efficient in the subsequent case as demonstrated by longer task completion times (r=−0.521; P=.007), higher numbers of mouse clicks (r=−0.562; P=.003), and more EHR screen visits (r=−0.486; P=.01). Visual strain and fatigue studies have focused on individual-level adjustments for accommodation at near distances, with lack of energy, physical discomfort, and sleepiness, were statistically significantly higher as functions of the length of session [13,14]. Thus, shifts at the workplace and organizational levels may be required for the overall workflow.

Approximately 45.8% (3338/7279) of the physicians worked for >60 hours per week compared with 10% (3442/34,420) of US workers in other fields [51]. Studies have determined that physicians spend 4 to 6 hours on EHR and desk work during the day and another 1 to 2 hours after work, often for clerical and administrative tasks (eg, documentation, order entry, billing, coding, and system security) [52]. Studies found that US providers compared with others spent substantially more time actively using the EHR (mean time 90.2 minutes vs 59.1 minutes; P<.001), including making notes, orders, in-basket messages, and clinical review [45]. They also composed more automated note text than their non-US counterparts (270/348, 77.5% vs 14/23, 61%; P<.001) and received statistically significantly more messages per day (33.8 vs 12.8; P<.001). Furthermore, US clinicians used the EHR for a longer time after hours, logging in 26.5 minutes per day versus 19.5 minutes per day for non-US clinicians (P=.01). These results persisted after controlling for organizational characteristics, including structure, type, size, and daily patient volume. The most important 3 factors that separate the ideal order sets from the rest are patient safety, efficiency, and user satisfaction. Scientific evidence, workflow, ordering efficiency, and user satisfaction reduce mouse clicks and unproductive thinking times [53].

Implementation studies of usability suggest that there are multiple opportunities to improve the use of EHRs across professions, particularly in relation to usability [47,51-53]. A survey on health information technology (IT) for pharmacy practice showed that some EHRs may also introduce new error types (eg, excessive alerts can lead to fatigue, so much so that providers can inadvertently ignore scanner barcode indicators of drug mismatches and erroneously identify drugs) [46].

Provider Perceptions and Experiences With Technology in Health Care Not Specific to Fatigue

A total of 14 studies explored provider experiences or perceptions about technology that may apply to fatigue but did not directly investigate it. These studies focused on EHR and videos (6/14, 43%); combinations of video display terminals (VDTs), computers, and phones (6/14, 43%); smartphones or PDA (1/14, 7%), or EHR alone (1/14, 7%; Multimedia Appendix 2 [6,56,61-72]). Methods were heterogeneous with surveys, semistructured interviews, qualitative methods, and comparison groups (eg, video vs in-person or other). The foci of the assessment were behavioral (9/14, 64%), emotional (9/14, 64%), cognitive (10/14, 71%), and physical impact (1/14, 7%) and the assessment was at the individual (12/14, 86%), clinic (6/14, 43%), hospital (6/14, 43%), and system or organizational (6/14, 43%) levels.

Studies that focused on EHR time log data for physicians [61,71] found substantial time of EHR use (eg, 5.9 hours of a 11.4-hour workday in a hospital, 4.5 hours during clinics hours, and 1.4 hours after clinic). Of time spent in the EHR, clerical and administrative tasks such as documentation, order entry, billing, and coding accounted for 44.2%, inbox management for 23.7%, and additional time communicating with patients, refilling prescriptions, or reviewing test results each day [61,71]. User ratings were high on data quality, accuracy, and processing [62] but low for satisfaction with clerical tasks [6]. Interrupted tasks require more time and result in more errors, stress, and frustration [72], and qualitative interviews and focus groups suggested more focus on usability, usefulness, training, and support [65,68]. There were differences among generations regarding adaptability, perceived benefits and drawbacks, and perceptions of other generations' abilities to adapt.

A Continuum From Health to Fatigue to Burnout

Qualitative analysis of the literature suggests a continuum from health to risk of fatigue to manifestations of digital burnout (Figure 2). This was stratified by clinical care, technology, routine, and social, interpersonal and professional dimensions. Related to care, providers vary in how aligned technology is with goals, how *therapeutic* or enjoyable it is for them (and not just patients), and other rewards. Organizations play a substantial role in selection and implementation of technology for clinical workflow, workload, and remuneration, which often predetermine routines. Provider input is sometimes solicited. When organizations use user-centered design or lean processes, user satisfaction and the fit of goals, methods, and routines may be much higher [59,60], avoiding gaps between the system and provider perspectives. Fatigue often manifests in social and interpersonal contexts, with taxing fatigue overtaking effectiveness and engagement, resulting in burnout with exhaustion, cynicism, and feelings of ineffectiveness [6,24,25].

Organizational Responses Based on Provider Experiences and Human Factors Related to Technology

A qualitative analysis suggested multiple opportunities for regulators, policy makers, EHR developers, payers, health system leadership, and users to collectively improve the use of EHRs and other technologies (Multimedia Appendix 3) [47,52]. It summarizes human factors and technology in health care: organizational responses for prevention and adjustment of workflow, as organized in terms of evidence and findings, manifestations and analysis of technology problems, and individual user versus organizational adjustments being made. A change management process for workflow and administration [73] requires training, supervision, and evaluation to adjust

XSL•FO
RenderX

competencies or skills, improve quality and performance indicators, and reallocate resources in health care [2,74].

Business, Occupational Health, and Well-being Literature to Contextualize Technology-Based Fatigue

Overview

From the 1980s to 2021, there has been a shift in the perception of the origin of technology problems in business, occupational health, and other area [46,75]. Earlier perceptions attributed problems of production, efficiency, and outputs as being related to ergonomic, mechanical, workplace, and other factors for individuals, cohorts (eg, VDT employees), and systems. More recently, gaining input from users of technology is central to the design of the workplace to minimize and prevent problems.

VDT Studies

Findings from VDT studies of occupational hygiene (eg, climate, lighting, and electrostatic conditions) and ophthalmologic dimensions appear to be quite pertinent to video, EHR, and psychosocial work commonly associated with *technology fatigue* [9,10,76]. Job stress has been found to be higher; quality of life has been found to be lower; and visual strain, discomfort, and fatigue has been found to be higher in VDT workers than in non-VDT workers [77], and combined, interactive communication causes more discomfort than data entry or acquisition. Postural risk factors and job strain in the environment seem to increase musculoskeletal symptoms for those with >7 hours of VDT use per day, but ergonomic interventions improve function [57,78].

Displays and workflow interventions have been successful in many respects. A 15-minute work period with microbreaks [79] and physically large displays help improve employee performance. A 15° rather than 40° video display curvature (display curvature impacts effort to visualize displayed text) [80] and a case manuscript and luminance ratio of 3 (used for the useful contrast of a display) also help users' performance [9-14]. Coworking spaces are an alternative to home offices because professional isolation negatively affects job performance, and employees with inhibitory deficits (eg, prone to distraction) and poorer boundaries (ie, limited psychological detachment) experience more stress [58,81].

Studies Assessing Fatigue and Burnout

There are overlaps and differences between burnout and prolonged fatigue [82]. Fatigue plays a central role in the development of burnout (ie, medical) and prolonged fatigue (ie, Psychological), with the former conceptualized as a work-related condition and prolonged fatigue as a general condition. Burnout manifests as exhaustion (physical and emotional), cynicism and detachment from the job and others, and a sense of ineffectiveness and lack of accomplishment [24,25]. Low job dissatisfaction is associated with low organizational commitment, absenteeism, intention to leave the job, turnover, lower productivity, and impaired quality of work. Those who experience burnout also disrupt job tasks and display greater interpersonal aggression [24,25]. The Areas of Worklife model considers workload, sense of control, reward, community, fairness, and values as important to burnout [75]. Rewards and recognition provide opportunities for intrinsic satisfaction and self-efficacy. A good community provides social support, trust, effective means of working out disagreement, and better job engagement. The job demands-resources model posits that burnout is related to the experience of incessant job demands and inadequate resources, whereas the conservation of resources model follows basic motivational theory in assuming that burnout arises because of persistent threats to available resources.

Conceptualization of fatigue and burnout may also be organized according to engagement and job stress [24,25,83]. Engagement is considered a state of high energy, strong involvement, and a sense of efficacy. It is a persistent, positive, affective-motivational state of fulfillment characterized by the 3 components of vigor, dedication, and absorption. Engagement is considered an independent and distinct concept, which is not the opposite of burnout, although is negatively related to it. Interventions at the *individual level* may involve the following: (1) changing work patterns, (2) developing coping skills, (3) obtaining social support, (4) using relaxation strategies, (5) promoting good health and fitness, and (6) developing a better self-understanding. At the *workplace or organizational level*, this may mean the following: (1) redesigning job tasks, (2) improving recognition of notable work by both teams and individuals, and (3) developing more fair and equitable policies.

Guidelines for Providers, Systems, and Organizations in Health Care for Use of Technology and Well-being

A shift to a culture of well-being with technology use requires the evaluation, implementation, and monitoring of individual, workplace, workflow, and institutional strategies (Multimedia Appendix 4 [2,6,7,23-25,28-30,37,44,49-53,56,74,83-85]). If the link between technology and fatigue is poorly recognized, changes in workflow processes and policies may not be carried out until the provider's well-being is already at risk owing to burnout [49]. Guidelines for health care, well-being, and the use of technology (eg, EHRs) to avoid burnout were found throughout these studies and summarized as well; however, these need to evaluate fatigue.

A shift to a culture of well-being that incorporates technology will require adaptations and quality improvement in the areas of technology, physical environment, occupational health, and specific evaluations and interventions. Therefore, objective measures to evaluate, promote, and enhance well-being are required. At a minimum, consideration is needed for the cognitive, behavioral, emotional, and physical impact of workflows. Such consideration is needed at the individual, clinic, hospital, and system or organizational levels. This could include adjustments in information systems (IS) and IT, use of lean methods and emphasis on interprofessional education efforts with technology team-based care from the Institute of Healthcare Improvement and Agency for Healthcare Research and Quality [2]. More specifically, methods are needed to evaluate clinical workflows, promote provider competencies with technology and self-care and implement institutional competencies for technologies. Deliberate, sustained, and comprehensive efforts by the organization are often inexpensive, reduce burnout, and promote engagement [24,25,49-51].

XSL•FO
RenderX

Health care provider well-being and health appear to be related to the technological integration of video, EHR, and mobile health over time (Figure 3). This figure was created based on time points related to the following: (1) the release of new technologies into the marketplace; (2) the introduction (or in some cases integration) of technologies into workflow for health care systems, which was generalized, as some systems integrated sooner and others later, and private practice providers were likely quite heterogeneously adapting; and (3) content and discourse analysis to thematically capture concept area terminology surfacing in the literature related to technology (eg, burnout has been identified much earlier, but fatigue and technology have surfaced in recent years). Organizational efforts and resultant outcomes for well-being may be stratified from high to low based on individual, system, and organizational contributions, as follows: (1) high—substantial institutional support to seek user input, optimize clinical skills, and monitor resilience and well-being; (2) mid—good but limited institutional support to include user input, which improves some workflow processes but not systematically; and (3) low—minimal institutional support with expectations that users learn, adapt workflow, and maintain well-being individually.

Discussion

Principal Findings

Studies related to the implementation and evaluation of technology are increasingly sophisticated and provide a starting place despite varying widely in duration, approaches, methods, and quality of measures. The 12 studies that met all the inclusion criteria for technology, health care, and fatigue studied the behavioral, emotional, cognitive, and physical impact of workflow at individual, hospital, and system or organizational levels (Multimedia Appendix 1) [8,13,14,45-53]. Video and EHR use is associated with fatigue with physical eye fatigue, neck pain, stress, and tiredness; behavioral impact related to additional effort owing to barriers, trouble with engagement, emotional wear and tear and exhaustion, cognitive inattention, effort, expecting problems, multitasking and workload, and emotional experiences such as anger, irritability, stress, and concern about well-being. An additional 14 studies that evaluated the behavioral, emotional, and cognitive impact of using technology without focusing on fatigue found high user ratings on data quality, accuracy, and processing but low satisfaction with clerical tasks, the effort required in work and interruptions costing time and resulting in more errors, stress, and frustration (Multimedia Appendix 2) [6,56,61-72]. Other problems contributing to fatigue may include the addition of workflow steps before and after clinical care is provided, often requiring sustained periods of technology use. Few studies have discussed the physical environment, occupational health approaches, mobile, telework, lean, process improvement, human factors, and user design approaches to workflows. A qualitative analysis of the literature suggests a continuum from health to risk of fatigue to manifestations of digital burnout, which provides an outline of organization approaches to human factors and technology in health care (Figure 2). Although business, occupational health, and well-being literature did not study technology fatigue and burnout, findings from the literature help contextualize technology-based fatigue and modern approaches they use such as lean, process improvement, occupational health, design studios, and implementation science that could be helpful in health care at individual, clinic, hospital, and system or organizational levels. Few studies were found to contextually evaluate differences according to health professions and practice contexts.

Beneficial changes in workplace culture, focus on well-being, and prevention of burnout from other fields are beginning to be used in health care [29,49,86], but an accurate evaluation of the problems is just beginning. Areas of specific focus include clinical care, human factors, training, professional development, workflow, and administration factors (Multimedia Appendix 3). The conceptualization of *burnout* is undergoing change, with a shift toward the recognition of burnout as an occupational phenomenon rather than solely as an individual medical disease (eg, depression) per the World Health Organization [87]. Thus, deployment of health care and administrative resources should move beyond the individual (eg, Family and Medical Leave Act) and look at structural, educational, cultural, and social factors.

To begin to address challenges in health care related to fatigue and burnout, including those associated with technology, a substantial collaborative effort is needed from health system leadership, organizational researchers, IT and IS specialists, and potentially the government [2,3,28,29], as changes in financing, reimbursement, and regulatory processes may need adjustment. An overall approach requires implementation, evaluation, and monitoring of individual, workplace, workflow, and institutional strategies (Multimedia Appendix 4). Financial support resources (eg, counseling, retirement planning, and college planning for children) can reduce competing demands for time and address personal and career concerns [51]. To transform organizational culture, *wrap-around* support for providers, not just patients, may be needed, as suggested by the Quadruple Aim. In the business culture of successful companies, such as Cirque du Soleil, L'Oreal Paris, and Nintendo, the tetrad foci of research, production, marketing, and finance have been expanded to a pentad by integrating technology rather than appending it [18,38,88]. An IT business–medicine understanding or conceptual framework has likewise been suggested with individual and institutional competencies [2,74] based on IT architecture [84].

A structural and functional redesign of systems would emphasize evaluation, effectiveness, implementation, and application of process improvement [59,60,85,89]. It includes approaches to causal questions using cross-sectional and longitudinal dimensions, multilevel foci, and objective measures for clinical (engagement, meaningfulness of tasks, process and quality measures, and clinical and safety outcomes), human factors (workload, rewards, fitness, needs, and well-being), training, professional development, and administrative (value alignment, productivity, IS, strategic planning, resources, and participative management) workflows. A 360° perspective with qualitative approaches could be useful to collect input from providers on what makes care *therapeutic*, enjoyable and easy to provide, and promotes their well-being and performance as individuals, team members, and leaders. In time, continuous data collection

XSL•FO
RenderX

and analytics could support clinical decision-making for patient quality, workforce satisfaction, and system outcomes, creating an organizational culture of well-being, compassion in care, and prevention of fatigue and burnout in all employees, including providers [2,3,90]. Human factors engineering and usability assessment has a rich set of scientific methods, a strong evidence base, and is widely applied effectively in other industries [84,85].

Limitations

This scoping review has some limitations. First, there were fewer findings than we expected using our inclusion and exclusion criteria, despite a broad scope, to find the relationships between health care, fatigue, and synchronous (video, telephone, and informatic systems) and asynchronous (wearable sensors and mobile health devices) technologies. Second, only 1 author reviewed the titles and abstracts. Third, the entire search was described but not saved and consolidated as an appendix for reviewers; reresearch findings of the 2 main databases were included as an appendix for reviewers. Although the terms, databases, and dates are a guide to other researchers, this omission does not enable others to simulate the approach. Fourth, given the small sample sizes, heterogeneous methods, and variable study duration, the team was unable to apply a systematic quality evaluation system or draw conclusions using a quantitative meta-analysis. Cross-sectional studies of associations with multiple factors in applied rather than controlled settings have limitations. Fifth, the stratification of behavioral, cognitive, emotional, and physical domains of impact, although heuristically helpful, could have been operationalized more rigorously. Similarly, *workplace* at the individual, clinic, hospital, and system or organizational levels may need better definitions. Sixth, the review does not cover all potentially relevant well-being, burnout, and stress dimensions of the workplace, nor does it cover research on the physical environment, occupational health, or mobile, virtual or telework workflows. Seventh, broader input for consensus across organizations could have been helpful, and a qualitative, small group interview approach with experts using a semistructured guide could have discovered more information.

Conclusions

Health care delivery and systems are increasingly incorporating technology but need to evaluate its impact in accordance with the Quadruple Aim to support providers. Approaches with causal questions and longitudinal implementation research could benefit from a multilevel approach with objective measures for clinical and human factors, training, professional development, and administrative workflows. If done well, technology integration could further population-centered health and effectiveness of service delivery, although the redesign of financing, reimbursement, regulatory, and other changes may be necessary. Integration of health care quality outcomes with those for technology and well-being is suggested and requires institutional strategies and competencies. Otherwise, continued advances in the use of technology may inadvertently worsen provider workload burden, fatigue, and burnout.

Acknowledgments

The authors would like to thank the American Telemedicine Association; Telemental Health Interest Group; Coalition for Technology in Behavioral Science; Department of Psychiatry and Behavioral Sciences, University of California; Davis School of Medicine; Veteran Affairs Northern California Health Care System; and Mental Health Service Office of Connected Care, Department of Veterans Affairs for their support.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Studies at the intersection of technology, health care, and fatigue.
[DOCX File , 36 KB - jmir_v24i5e34451_app1.docx]

Multimedia Appendix 2
Studies of the reviews focused on technology-related experiences in health care for clinicians aside from fatigue.
[DOCX File , 35 KB - jmir_v24i5e34451_app2.docx]

Multimedia Appendix 3
Human factors and technology in health care: organizational responses for prevention and adjustment of workflow.
[DOCX File , 41 KB - jmir_v24i5e34451_app3.docx]

Multimedia Appendix 4
Guidelines for clinicians, systems, and organizations in health care for technology use and well-being.
[DOCX File , 22 KB - jmir_v24i5e34451_app4.docx]

References

XSL•FO
RenderX

1. Anderson J, Rainie L, Vogels EA. Experts Say the 'New Normal' in 2025 Will Be Far More Tech-Driven, Presenting More Big Challenges. Pew Research Center. 2021 Feb 18. URL: https://www.pewresearch.org/internet/2021/02/18/experts-say-the-new-normal-in-2025-will-be-far-more-tech-driven-presenting-more-big-challenges/ [accessed 2022-02-16]
2. Hilty DM, Torous J, Parish MB, Chan SR, Xiong G, Scher L, et al. A literature review comparing clinicians' approaches and skills to in-person, synchronous, and asynchronous care: moving toward competencies to ensure quality care. Telemed J E Health 2021 Apr;27(4):356-373. [doi: 10.1089/tmj.2020.0054] [Medline: 32412882]
3. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
4. Pattyn N, Van Cutsem J, Dessy E, Mairesse O. Bridging exercise science, cognitive psychology, and medical practice: is "cognitive fatigue" a remake of "the emperor's new clothes"? Front Psychol 2018 Sep 10;9:1246 [FREE Full text] [doi: 10.3389/fpsyg.2018.01246] [Medline: 30250436]
5. Montgomery A, Panagopoulou E, Esmail A, Richards T, Maslach C. Burnout in healthcare: the case for organisational change. BMJ 2019 Jul 30;366:l4774. [doi: 10.1136/bmj.l4774] [Medline: 31362957]
6. Shanafelt TD, Dyrbye LN, Sinsky C, Hasan O, Satele D, Sloan J, et al. Relationship between clerical burden and characteristics of the electronic environment with physician burnout and professional satisfaction. Mayo Clin Proc 2016 Jul;91(7):836-848. [doi: 10.1016/j.mayocp.2016.05.007] [Medline: 27313121]
7. Nguyen OT, Jenkins NJ, Khanna N, Shah S, Gartland AJ, Turner K, et al. A systematic review of contributing factors of and solutions to electronic health record-related impacts on physician well-being. J Am Med Inform Assoc 2021 Apr 23;28(5):974-984 [FREE Full text] [doi: 10.1093/jamia/ocaa339] [Medline: 33517382]
8. 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] [Medline: 32515799]
9. Knave BG, Wibom RI, Bergqvist UO, Carlsson LL, Levin MI, Nylén PR. Work with video display terminals among office employees. II. Physical exposure factors. Scand J Work Environ Health 1985 Dec;11(6):467-474 [FREE Full text] [doi: 10.5271/sjweh.2199] [Medline: 4095525]
10. Böös SR, Calissendorff BM, Knave BG, Nyman KG, Voss M. Work with video display terminals among office employees. III. Ophthalmologic factors. Scand J Work Environ Health 1985 Dec;11(6):475-481 [FREE Full text] [doi: 10.5271/sjweh.2207] [Medline: 3912987]
11. Travers PH, Stanton BA. Office workers and video display terminals: physical, psychological and ergonomic factors. AAOHN J 2002 Nov;50(11):489-493. [doi: 10.1177/216507990205001102] [Medline: 12465203]
12. Mocci F, Serra A, Corrias GA. Psychological factors and visual fatigue in working with video display terminals. Occup Environ Med 2001 Apr;58(4):267-271 [FREE Full text] [doi: 10.1136/oem.58.4.267] [Medline: 11245744]
13. Krupinski EA, Berbaum KS. Measurement of visual strain in radiologists. Acad Radiol 2009 Aug;16(8):947-950 [FREE Full text] [doi: 10.1016/j.acra.2009.02.008] [Medline: 19406673]
14. Krupinski EA, Berbaum KS, Caldwell RT, Schartz KM, Kim J. Long radiology workdays reduce detection and accommodation accuracy. J Am Coll Radiol 2010 Sep;7(9):698-704 [FREE Full text] [doi: 10.1016/j.jacr.2010.03.004] [Medline: 20816631]
15. Dawson D, Ian Noy Y, Härmä M, Akerstedt T, Belenky G. Modelling fatigue and the use of fatigue models in work settings. Accid Anal Prev 2011 Mar;43(2):549-564. [doi: 10.1016/j.aap.2009.12.030] [Medline: 21130216]
16. Caldwell JA, Caldwell JL, Thompson LA, Lieberman HR. Fatigue and its management in the workplace. Neurosci Biobehav Rev 2019 Jan;96:272-289 [FREE Full text] [doi: 10.1016/j.neubiorev.2018.10.024] [Medline: 30391406]
17. Lester CA, Tu L, Ding Y, Flynn AJ. Detecting potential medication selection errors during outpatient pharmacy processing of electronic prescriptions with the RxNorm application programming interface: retrospective observational cohort study. JMIR Med Inform 2020 Mar 11;8(3):e16073 [FREE Full text] [doi: 10.2196/16073] [Medline: 32044760]
18. Rivard S, Pinsonneault A, Croteau AM. Information technology at Cirque du Soleil: Looking back, moving forward. Thirty Second International Conference on Information Systems. 2012. URL: https://core.ac.uk/download/pdf/301352433.pdf [accessed 2022-04-25]
19. Maricu oiu LP, Sulea C, Iancu A. Work engagement or burnout: which comes first? A meta-analysis of longitudinal evidence. Burnout Res 2017 Jun;5:35-43 [FREE Full text] [doi: 10.1016/j.burn.2017.05.001]
20. Mäkikangas A, Leiter MP, Kinnunen U, Feldt T. Profiling development of burnout over eight years: relation with job demands and resources. Eur J Work Organ Psychol 2021;30(5):720-731 [FREE Full text] [doi: 10.1080/1359432x.2020.1790651]
21. Schneider A, Weigl M. Associations between psychosocial work factors and provider mental well-being in emergency departments: a systematic review. PLoS One 2018 Jun 4;13(6):e0197375 [FREE Full text] [doi: 10.1371/journal.pone.0197375] [Medline: 29864128]
22. Sakuraya A, Imamura K, Watanabe K, Asai Y, Ando E, Eguchi H, et al. What kind of intervention is effective for improving subjective well-being among workers? A systematic review and meta-analysis of randomized controlled trials. Front Psychol 2020 Nov 13;11:528656 [FREE Full text] [doi: 10.3389/fpsyg.2020.528656] [Medline: 33281653]

XSL•FO
RenderX

23. 7 steps to prevent burnout in your practice. American Medical Association. 2015 Aug 11. URL: https://www.ama-assn.org/practice-management/physician-health/7-steps-prevent-burnout-your-practice [accessed 2022-02-16]
24. Maslach C, Leiter MP. Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry 2016 Jun;15(2):103-111 [FREE Full text] [doi: 10.1002/wps.20311] [Medline: 27265691]
25. Maslach C, Leiter MP. New insights into burnout and health care: strategies for improving civility and alleviating burnout. Med Teach 2017 Feb;39(2):160-163. [doi: 10.1080/0142159X.2016.1248918] [Medline: 27841065]
26. West CP, Dyrbye LN, Erwin PJ, Shanafelt TD. Interventions to prevent and reduce physician burnout: a systematic review and meta-analysis. Lancet 2016 Nov 05;388(10057):2272-2281. [doi: 10.1016/S0140-6736(16)31279-X] [Medline: 27692469]
27. Stehman C, Testo Z, Gershaw R, Kellogg AR. Burnout, drop out, suicide: physician loss in emergency medicine, Part I. West J Emerg Med 2019 May;20(3):485-494 [FREE Full text] [doi: 10.5811/westjem.2019.4.40970] [Medline: 31123550]
28. Swensen SJ, Shanafelt T. An organizational framework to reduce professional burnout and bring back joy in practice. Jt Comm J Qual Patient Saf 2017 Jun;43(6):308-313. [doi: 10.1016/j.jcjq.2017.01.007] [Medline: 28528625]
29. Shanafelt TD, Noseworthy JH. Executive leadership and physician well-being: nine organizational strategies to promote engagement and reduce burnout. Mayo Clin Proc 2017 Jan;92(1):129-146. [doi: 10.1016/j.mayocp.2016.10.004] [Medline: 27871627]
30. Downing NL, Bates DW, Longhurst CA. Physician burnout in the electronic health record era: are we ignoring the real cause? Ann Intern Med 2018 Jul 03;169(1):50-51. [doi: 10.7326/M18-0139] [Medline: 29801050]
31. Gleichgerrcht E, Decety J. Empathy in clinical practice: how individual dispositions, gender, and experience moderate empathic concern, burnout, and emotional distress in physicians. PLoS One 2013 Apr 19;8(4):e61526 [FREE Full text] [doi: 10.1371/journal.pone.0061526] [Medline: 23620760]
32. Fothergill A, Edwards D, Burnard P. Stress, burnout, coping and stress management in psychiatrists: findings from a systematic review. Int J Soc Psychiatry 2004 Mar;50(1):54-65. [doi: 10.1177/0020764004040953] [Medline: 15143847]
33. Rössler W. Stress, burnout, and job dissatisfaction in mental health workers. Eur Arch Psychiatry Clin Neurosci 2012 Nov;262 Suppl 2:S65-S69. [doi: 10.1007/s00406-012-0353-4] [Medline: 22926058]
34. O'Connor K, Muller Neff D, Pitman S. Burnout in mental health professionals: a systematic review and meta-analysis of prevalence and determinants. Eur Psychiatry 2018 Sep;53:74-99. [doi: 10.1016/j.eurpsy.2018.06.003] [Medline: 29957371]
35. Panagioti M, Panagopoulou E, Bower P, Lewith G, Kontopantelis E, Chew-Graham C, et al. Controlled interventions to reduce burnout in physicians: a systematic review and meta-analysis. JAMA Intern Med 2017 Feb 01;177(2):195-205. [doi: 10.1001/jamainternmed.2016.7674] [Medline: 27918798]
36. Van Steenbergen EF, van der Ven C, Peeters MC, Taris TW. Transitioning towards new ways of working: do job demands, job resources, burnout, and engagement change? Psychol Rep 2018 Aug;121(4):736-766 [FREE Full text] [doi: 10.1177/0033294117740134] [Medline: 29298562]
37. Hilty DM, Crawford A, Teshima J, Chan S, Sunderji N, Yellowlees PM, et al. A framework for telepsychiatric training and e-health: competency-based education, evaluation and implications. Int Rev Psychiatry 2015;27(6):569-592. [doi: 10.3109/09540261.2015.1091292] [Medline: 26540642]
38. Hilty DM, Luo J, Giron E, Ko DG. A shared information technology-business-health model: lessons for healthcare leaders on integrating technology from investment. Psychol Cogn Sci Open J 2021 Jul 06;7(1):1-18. [doi: 10.17140/pcsoj-7-159]
39. 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]
40. 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] [Medline: 20854677]
41. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
42. Crowe M, Inder M, Porter R. Conducting qualitative research in mental health: thematic and content analyses. Aust N Z J Psychiatry 2015 Jul;49(7):616-623. [doi: 10.1177/0004867415582053] [Medline: 25900973]
43. Nijrolder I, van der Horst H, van der Windt D. Prognosis of fatigue. A systematic review. J Psychosom Res 2008 Apr;64(4):335-349. [doi: 10.1016/j.jpsychores.2007.11.001] [Medline: 18374732]
44. de Villiers MR, de Villiers PJ, Kent AP. The Delphi technique in health sciences education research. Med Teach 2005 Nov;27(7):639-643. [doi: 10.1080/13611260500069947] [Medline: 16332558]
45. Holmgren AJ, Downing NL, Bates DW, Shanafelt TD, Milstein A, Sharp CD, et al. Assessment of electronic health record use between us and non-US health systems. JAMA Intern Med 2021 Feb 01;181(2):251-259 [FREE Full text] [doi: 10.1001/jamainternmed.2020.7071] [Medline: 33315048]
46. Galt KA, Fuji KT, Kaufman TK, Shah SR. Health information technology use and patient safety: study of pharmacists in Nebraska. Pharmacy (Basel) 2019 Jan 10;7(1):7 [FREE Full text] [doi: 10.3390/pharmacy7010007] [Medline: 30634591]
47. Lopez KD, Fahey L. Advocating for greater usability in clinical technologies: the role of the practicing nurse. Crit Care Nurs Clin North Am 2018 Jun;30(2):247-257. [doi: 10.1016/j.cnc.2018.02.007] [Medline: 29724443]

XSL•FO
RenderX

48. Mylod DE. One Way to Prevent Clinician Burnout. Harvard Business Review. 2017 Oct 12. URL: https://hbr.org/2017/10/one-way-to-prevent-clinician-burnout [accessed 2022-02-16]
49. Nakagawa K, Yellowlees P. Inter-generational effects of technology: why millennial physicians may be less at risk for burnout than baby boomers. Curr Psychiatry Rep 2020 Jul 13;22(9):45 [FREE Full text] [doi: 10.1007/s11920-020-01171-2] [Medline: 32661817]
50. National Academy of Medicine, National Academies of Sciences, Engineering, and Medicine. Taking Action Against Clinician Burnout: A Systems Approach to Professional Well-Being. National Academy of Medicine. 2019. URL: https://tinyurl.com/2p9bdef9 [accessed 2022-02-16]
51. Shanafelt TD, Lightner DJ, Conley CR, Petrou SP, Richardson JW, Schroeder PJ, et al. An organization model to assist individual physicians, scientists, and senior health care administrators with personal and professional needs. Mayo Clin Proc 2017 Nov;92(11):1688-1696. [doi: 10.1016/j.mayocp.2017.08.020] [Medline: 29101937]
52. Tutty MA, Carlasare LE, Lloyd S, Sinsky CA. The complex case of EHRs: examining the factors impacting the EHR user experience. J Am Med Inform Assoc 2019 Jul 01;26(7):673-677 [FREE Full text] [doi: 10.1093/jamia/ocz021] [Medline: 30938754]
53. Zhang Y, Padman R, Levin JE. Reducing provider cognitive workload in CPOE use: optimizing order sets. Stud Health Technol Inform 2013;192:734-738. [Medline: 23920654]
54. Braithwaite J, Churruca K, Long JC, Ellis LA, Herkes J. When complexity science meets implementation science: a theoretical and empirical analysis of systems change. BMC Med 2018 Apr 30;16(1):63 [FREE Full text] [doi: 10.1186/s12916-018-1057-z] [Medline: 29706132]
55. Keyworth C, Hart J, Armitage CJ, Tully MP. What maximizes the effectiveness and implementation of technology-based interventions to support healthcare professional practice? A systematic literature review. BMC Med Inform Decis Mak 2018 Nov 07;18(1):93 [FREE Full text] [doi: 10.1186/s12911-018-0661-3] [Medline: 30404638]
56. Middleton B, Bloomrosen M, Dente MA, Hashmat B, Koppel R, Overhage JM, American Medical Informatics Association. Enhancing patient safety and quality of care by improving the usability of electronic health record systems: recommendations from AMIA. J Am Med Inform Assoc 2013 Jun;20(e1):e2-e8 [FREE Full text] [doi: 10.1136/amiajnl-2012-001458] [Medline: 23355463]
57. Karwowski W, Eberts R, Salvendy G, Noland S. The effects of computer interface design on human postural dynamics. Ergonomics 1994 Apr;37(4):703-724. [doi: 10.1080/00140139408963684] [Medline: 8187753]
58. Robelski S, Keller H, Harth V, Mache S. Coworking spaces: the better home office? A psychosocial and health-related perspective on an emerging work environment. Int J Environ Res Public Health 2019 Jul 04;16(13):2379 [FREE Full text] [doi: 10.3390/ijerph16132379] [Medline: 31277440]
59. Aij KH, Rapsaniotis S. Leadership requirements for Lean versus servant leadership in health care: a systematic review of the literature. J Healthc Leadersh 2017 Jan 18;9:1-14 [FREE Full text] [doi: 10.2147/JHL.S120166] [Medline: 29355240]
60. Maijala R, Eloranta S, Reunanen T, Ikonen TS. Successful implementation of Lean as a managerial principle in health care: a conceptual analysis from systematic literature review. Int J Technol Assess Health Care 2018 Jan;34(2):134-146. [doi: 10.1017/S0266462318000193] [Medline: 29642955]
61. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan WJ, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. Ann Fam Med 2017 Sep;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
62. Campbell EM, Sittig DF, Guappone KP, Dykstra RH, Ash JS. Overdependence on technology: an unintended adverse consequence of computerized provider order entry. AMIA Annu Symp Proc 2007 Oct 11;2007:94-98 [FREE Full text] [Medline: 18693805]
63. Hilty DM, Marks S, Wegelin J, Callahan EJ, Nesbitt TS. A randomized, controlled trial of disease management modules, including telepsychiatric care, for depression in rural primary care. Psychiatry (Edgmont) 2007 Feb;4(2):58-65 [FREE Full text] [Medline: 20805900]
64. McAlearney AS, Schweikhart SB, Medow MA. Organizational and physician perspectives about facilitating handheld computer use in clinical practice: results of a cross-site qualitative study. J Am Med Inform Assoc 2005;12(5):568-575 [FREE Full text] [doi: 10.1197/jamia.M1816] [Medline: 15905482]
65. McAlearney AS, Hefner JL, Sieck CJ, Huerta TR. The journey through grief: insights from a qualitative study of electronic health record implementation. Health Serv Res 2015 Apr;50(2):462-488 [FREE Full text] [doi: 10.1111/1475-6773.12227] [Medline: 25219627]
66. Mirhafez SR, Movahedi A, Moghadam-Pasha A, Mohammadi G, Moeini V, Moradi Z, et al. Perceptions and practices related to clinical alarms. Nurs Forum 2019 Jul;54(3):369-375. [doi: 10.1111/nuf.12338] [Medline: 30838672]
67. Nakagawa K, Yellowlees PM. University of California Technology Wellness Index: a physician-centered framework to assess technologies' impact on physician well-being. Psychiatr Clin North Am 2019 Dec;42(4):669-681. [doi: 10.1016/j.psc.2019.08.005] [Medline: 31672216]
68. Nimjee T, Miller E, Solomon S. Exploring generational differences in physicians' perspectives on the proliferation of technology within the medical field: a narrative study. Healthc Q 2020 May;23(SP):53-59. [doi: 10.12927/hcq.2020.26172] [Medline: 32333748]

XSL•FO
RenderX

69. Paterick ZR, Patel NJ, Paterick TE. Unintended consequences of the electronic medical record on physicians in training and their mentors. Postgrad Med J 2018 Nov;94(1117):659-661. [doi: 10.1136/postgradmedj-2018-135849] [Medline: 30554171]
70. Sockalingam S, Clarkin C, Serhal E, Pereira C, Crawford A. Responding to health care professionals' mental health needs during COVID-19 through the rapid implementation of project ECHO. J Contin Educ Health Prof 2020;40(3):211-214. [doi: 10.1097/CEH.0000000000000311] [Medline: 32658015]
71. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. Health Aff (Millwood) 2017 Apr 01;36(4):655-662 [FREE Full text] [doi: 10.1377/hlthaff.2016.0811] [Medline: 28373331]
72. Westbrook JI, Coiera E, Dunsmuir WT, Brown BM, Kelk N, Paoloni R, et al. The impact of interruptions on clinical task completion. Qual Saf Health Care 2010 Aug;19(4):284-289. [doi: 10.1136/qshc.2009.039255] [Medline: 20463369]
73. Christensen CM. Disruptive innovation: can health care learn from other industries? A conversation with Clayton M. Christensen. Interview by Mark D. Smith. Health Aff (Millwood) 2007;26(3):w288-w295. [doi: 10.1377/hlthaff.26.3.w288] [Medline: 17355981]
74. Hilty DM, Unützer J, Ko DG, Luo J, Worley LL, Yager J. Approaches for departments, schools, and health systems to better implement technologies used for clinical care and education. Acad Psychiatry 2019 Dec;43(6):611-616. [doi: 10.1007/s40596-019-01074-2] [Medline: 31172384]
75. Leiter MP, Maslach C. Areas of Worklife Survey: Manual and Sampler Set. 5th edition. Menlo Park, CA, USA: Mind Garden; 2011.
76. Steffy BD, Jones JW. The psychological impact of video display terminals on employees' well-being. Am J Health Promot 1989;4(2):101-107. [doi: 10.4278/0890-1171-4.2.101] [Medline: 10296285]
77. Turville KL, Psihogios JP, Ulmer TR, Mirka GA. The effects of video display terminal height on the operator: a comparison of the 15 degree and 40 degree recommendations. Appl Ergon 1998 Aug;29(4):239-246. [doi: 10.1016/s0003-6870(97)00048-3] [Medline: 9701537]
78. Aarås A, Horgen G, Ro O, Løken E, Mathiasen G, Bjørset HH, et al. The effect of an ergonomic intervention on musculoskeletal, psychosocial and visual strain of VDT data entry work: the Norwegian part of the international study. Int J Occup Saf Ergon 2005;11(1):25-47. [doi: 10.1080/10803548.2005.11076627] [Medline: 15794872]
79. Balci R, Aghazadeh F. The effect of work-rest schedules and type of task on the discomfort and performance of VDT users. Ergonomics 2003 Apr 15;46(5):455-465. [doi: 10.1080/0014013021000047557] [Medline: 12745696]
80. Park S, Kyung G, Choi D, Yi J, Lee S, Choi B, et al. Effects of display curvature and task duration on proofreading performance, visual discomfort, visual fatigue, mental workload, and user satisfaction. Appl Ergon 2019 Jul;78:26-36. [doi: 10.1016/j.apergo.2019.01.014] [Medline: 31046957]
81. Golden TD, Veiga JF, Dino RN. The impact of professional isolation on teleworker job performance and turnover intentions: does time spent teleworking, interacting face-to-face, or having access to communication-enhancing technology matter? J Appl Psychol 2008 Nov;93(6):1412-1421. [doi: 10.1037/a0012722] [Medline: 19025257]
82. Leone SS, Huibers MJ, Knottnerus JA, Kant IJ. Similarities, overlap and differences between burnout and prolonged fatigue in the working population. QJM 2007 Oct;100(10):617-627. [doi: 10.1093/qjmed/hcm073] [Medline: 17921196]
83. Molino M, Cortese CG, Ghislieri C. The promotion of technology acceptance and work engagement in industry 4.0: from personal resources to information and training. Int J Environ Res Public Health 2020 Apr 03;17(7):2438 [FREE Full text] [doi: 10.3390/ijerph17072438] [Medline: 32260142]
84. Ross JW. Creating a strategic IT architecture competency: Learning in stages. MIT Sloan School of Management. 2003 Jun 27. URL: http://dspace.mit.edu/handle/1721.1/3526 [accessed 2022-02-16]
85. Szalma JL. On the application of motivation theory to human factors/ergonomics: motivational design principles for human-technology interaction. Hum Factors 2014 Dec;56(8):1453-1471. [doi: 10.1177/0018720814553471] [Medline: 25509824]
86. Perlo J, Balik B, Swensen S, Landsman J, Feeley D. IHI Framework for Improving Joy in Work. IHI White Paper. Cambridge, MA, USA: Institute for Healthcare Improvement; 2017. URL: http://www.ihi.org/resources/Pages/IHIWhitePapers/Framework-Improving-Joy-in-Work.aspx [accessed 2022-02-16]
87. Burn-out an "occupational phenomenon": International Classification of Diseases. World Health Organization. 2019 May 28. URL: https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases [accessed 2022-02-16]
88. Ray G, Muhanna WA, Barney JB. Competing with IT: the role of shared IT-business understanding. Commun ACM 2007 Dec;50(12):87-91. [doi: 10.1145/1323688.1323700]
89. Pelayo S, Kaipio J, Section Editors for the IMIA Yearbook Section on Human Factors and Organizational Issues. Findings from the 2018 Yearbook Section on Human Factors and Organizational Issues. Yearb Med Inform 2018 Aug;27(1):79-82 [FREE Full text] [doi: 10.1055/s-0038-1667074] [Medline: 30157509]
90. Lown BA, Shin A, Jones RN. Can organizational leaders sustain compassionate, patient-centered care and mitigate burnout? J Healthc Manag 2019;64(6):398-412. [doi: 10.1097/JHM-D-18-00023] [Medline: 31725567]

XSL•FO
RenderX

Abbreviations

EHR: electronic health record
IS: information systems
IT: information technology
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
VDT: video display terminal

Edited by A Mavragani; submitted 01.11.21; peer-reviewed by K Hélène, T Ndabu, F Velayati, H Ayatollahi, PG Tian; comments to author 18.01.22; revised version received 20.02.22; accepted 18.04.22; published 25.05.22.

Please cite as:
Hilty DM, Armstrong CM, Smout SA, Crawford A, Maheu MM, Drude KP, Chan S, Yellowlees PM, Krupinski EA
Findings and Guidelines on Provider Technology, Fatigue, and Well-being: Scoping Review
J Med Internet Res 2022;24(5):e34451
URL: https://www.jmir.org/2022/5/e34451
doi:10.2196/34451
PMID:35612880

©Donald M Hilty, Christina M Armstrong, Shelby A Smout, Allison Crawford, Marlene M Maheu, Kenneth P Drude, Steven Chan, Peter M Yellowlees, Elizabeth A Krupinski. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2022. 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.

XSL•FO
RenderX

Review

Comparison of the ACASI Mode to Other Survey Modes in Sexual Behavior Surveys in Asia and Sub-Saharan Africa: Systematic Literature Review

Nang Nge Nge Phoo[1], MBBS, MPH; Roanna Lobo[1], BA, MSc, PhD; Daniel Vujcich[1], BA (Hons), LLB (Hons), MPhil, DPhil; Alison Reid[1], BSc (Hons), MSc, PhD

School of Population Health, Curtin University, Perth, Australia

Corresponding Author:
Nang Nge Nge Phoo, MBBS, MPH
School of Population Health
Curtin University
Kent Street
Bentley WA
Perth, 6102
Australia
Phone: 61 0414954410
Email: nangngenge.phoo@curtin.edu.au

Abstract

Background: Reliable data about sexual behaviors is fundamental in the prevention and control of HIV, hepatitis, and other sexually transmitted infections. Generally, sexual behaviors are regarded as a sociocultural taboo in Africa and Asia, and this results in biased sexual behavior survey data due to social desirability. Various modes of survey delivery, including audio computer-assisted self-interviews (ACASIs), have been investigated to improve data quality.

Objective: This study aimed to review studies that compared the ACASI mode to other survey modes in sexual behavior surveys in Asia and sub-Saharan Africa to ascertain the impact of survey mode on responses to sexual behavior questions.

Methods: A systematic literature review was conducted according to the Joanna Briggs Institute Manual for Evidence Synthesis. The review protocol was registered at PROSPERO (International Prospective Register of Systematic Reviews). Six databases were searched.

Results: A total of 21 papers were included. The face-to-face interview (FTFI) mode was the survey mode most frequently compared to the ACASI mode. Among the most commonly reported outcome variable groups, ACASI participants were more likely to report sexual behaviors, such as “forced sex,” “multiple partners,” “transactional sex,” and “ever had sex,” as compared to FTFI participants. In addition to the survey mode effect, other factors were found to have had an impact on data quality, for example, participant characteristics, social norms, study design, and data collection setting.

Conclusions: Use of ACASIs for administering sexual behavior surveys among populations in Asia and sub-Saharan Africa demonstrated higher reports for some sexual behaviors than the use of FTFIs. More studies that compare the ACASI mode to other survey modes would improve our understanding of the usefulness of ACASIs in sexual behavior surveys in these regions.

(J Med Internet Res 2022;24(5):e37356) doi:10.2196/37356

KEYWORDS

ACASI; survey mode; sexual behaviors; HIV; STI; hepatitis; blood-borne virus; Asia; sub-Saharan Africa; review

Introduction

While the annual number of new HIV cases decreased globally over the last decade from 2.1 million in 2010 to 1.5 million in 2020, there remained disparities across regions [1]. The HIV infection rate per 1000 uninfected population members was highest in the World Health Organization (WHO) African Region in 2020 (ie, 0.82 in Africa compared to 0.18 and lower in other regions) [1]. Similarly, in 2020, HIV-related deaths were higher in the WHO African Region (n=460,000) and in Southeast Asia (n=82,000) compared to other regions (n≤45,000) [1]. Furthermore, the prevalence rates of hepatitis B infection

XSL•FO
RenderX

and other sexually transmitted infections (STIs) were highest in the WHO African Region compared to all other regions [2-4].

Reliable data about sexual behaviors is fundamental in the prevention and control of HIV, hepatitis, and other STIs. Such data can help predict the transmission of STIs and inform strategies to prevent their transmission [5,6]. Development of intervention strategies requires exploration of the presence or frequency of each behavior and the risk and protective factors associated with each behavior [5,6]. Despite frequent research around sexual behaviors in Africa, this information was scarce for Asia [7].

Information on sexual behaviors is largely collected through self-reported knowledge, attitude, and practice surveys, for example, sexual behavior surveys in Thailand, Kenya, the United States, European countries, and Middle Eastern countries [8-11]. However, it is well recognized that the quality of self-reported data can be undermined in research about socially stigmatized behaviors [12-15]. Generally, "sexual behaviors" as a topic is a sensitive one and is regarded as taboo in Africa and Asia [16-19]. Consequently, attempts have been made to understand whether certain aspects of survey design can improve the quality of self-reported sexual behavior data. Previous research noted that the factors that impact on responses in sexual behavior surveys include the surveyed population, social context, respondent variables, an individual's choice, study design, data collection tools, and survey modes [20-22]. For instance, regarding survey modes, the involvement of interviewers can lead to a lack of privacy and increased social desirability bias, whereas self-administered surveys require minimum levels of literacy [12,15].

An audio computer-assisted self-interview (ACASI) is an electronic survey in which respondents listen to prerecorded survey questions, while reading the survey questions on desktops, laptops, or touch-screen devices, and enter their responses using a mouse, keyboard, color- or number-coded keypad, or touch screen [23]. Many studies have tested the use of ACASIs in capturing sensitive information with the hypotheses that noninvolvement of interviewers would enhance privacy and that the speech function would help respondents overcome any literacy issues, thus decreasing item nonresponse rates and increasing the prevalence of self-reported sexual experiences [24-27]. There have been a few reviews of studies comparing responses obtained from different survey modes, including ACASI and other computerized administration modes [24-27], with responses obtained from more traditional survey modes (eg, face-to-face interviews [FTFIs]). A review of 26 studies examining sexual behaviors of adolescents and adults in developing countries, in which 18 studies compared responses from ACASIs with those from other survey modes, found a higher rate of reporting of sensitive behaviors with computer-assisted interviews than with the other modes [24]. One of the papers included in that review reported that women were more likely to report anal sex in ACASIs (33%) than in FTFIs (24%; $P<.01$) [24,28]. A similar conclusion was reported by a meta-analysis of 48 studies, mostly from the United States and Europe [26], which compared the prevalence of self-reported sexual behaviors reported in computerized self-administration surveys with the prevalence in paper-and-pencil self-administration surveys regarding sensitive behaviors. A meta-analysis of the results from 15 survey studies about sexual behaviors and intravenous drug use compared reporting rates between face-to-face and non–face-to-face survey methods; in contrast to the previous meta-analysis, this one observed that the ACASI mode was not always associated with higher reporting of such behaviors, except for the studies conducted in Asia, urban areas, and among participants with secondary education and higher [25]. However, more than 60% of the studies assessed in these reviews were conducted in areas other than Africa or Asia, and ACASI was not the main survey mode of interest [24-26]. Therefore, there is a gap in knowledge about the usefulness of using the ACASI mode compared with other survey modes in sexual behavior surveys conducted in Africa and Asia.

Given the high incidence of HIV, hepatitis, and other STIs in the WHO African Region and in Asia, as well as scarce sexual behavior information in Asia, there is a need to identify a survey mode that best collects sensitive information around sexual behaviors for these populations. The aim of this study was to identify and review the studies conducted in Asia and sub-Saharan Africa that compared the ACASI mode to other survey modes in sexual behavior surveys in order to assess the impact of different survey modes on reporting of sexual behaviors.

Methods

Overview

A systematic literature review was conducted according to the Joanna Briggs Institute (JBI) Manual for Evidence Synthesis [29]. The review protocol was registered at PROSPERO (International Prospective Register of Systematic Reviews; registration No. CRD42020197237). A systematic search was conducted in six databases: Embase, MEDLINE, ProQuest Public Health Database, PsycInfo, Web of Science, and Global Health. The search commenced in June 2020 and was updated in August 2021. The search terms used were as follows: "ACASI," "audio computer," "survey mode," "HIV," "AIDS," "sex," "STI," "STD" (sexually transmitted disease), "syphilis," "chlamydia," "gonorrh*," "hepatitis," "blood-born*," "BBV" (blood-borne virus), and the names of the countries in Asia and sub-Saharan Africa. The search strategies used in the MEDLINE database are included in Figure S1 and Table S1 in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Population

The eligible population included adults and adolescents aged 15 years and older living in 36 countries in Southeast Asia, Northeast Asia, South Asia, and Central Asia, and in 48 countries in sub-Saharan Africa as listed by the World Bank in 2020 [30]. The population was limited to adults and mature young adults because this review was conducted as formative research for the original research which would have been carried out among adult populations.

XSL•FO
RenderX

Intervention

The intervention of interest was the use of ACASI in surveys about sexual behaviors and knowledge, attitudes, and practices around HIV, other STIs, and BBVs. The studies that focused on antiretroviral therapy and contraception use or adherence were not selected.

Comparators

The comparators to the ACASI mode were other survey modes with or without biological marker tests. When the ACASI mode was compared to biological markers only, the studies were excluded.

Outcomes

The outcome of interest was the difference in odds or prevalence of self-reported sexual behaviors between the ACASI mode and other survey modes. Moreover, the comparisons of discordance between the ACASI mode and biological marker results and that between other survey modes and biological marker results were also examined. Studies were also excluded when the outcome of interest was only the comparison in item response or refusal rates and user experience between the ACASI mode and other survey modes.

Resource Type, Publication Language, and Publication Period

Only peer-reviewed articles of primary studies published in English from 2000 to July 2021 were included.

Screening

The titles and abstracts of search results were screened against inclusion and exclusion criteria. The full text of papers with relevant titles and abstracts were retrieved and checked. The papers with relevant full text were included. The bibliographies of the included papers and the papers that cited the included papers were screened against the eligibility criteria. The search and screening results are presented in a 2020 PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (see Results section) [31].

Generic Study Characteristics and Delivery of the ACASI Mode

Data extraction was performed according to the JBI data extraction form for prevalence studies [29]. This included citation details, the generic study details, and description of main results. In addition, details relating to the focus of this review were extracted, including modes of survey administration, how the ACASI survey was delivered, and number of people or odds of reporting the presence of each sexual behavior by each survey mode. The generic study characteristics were summarized using tables, and a narrative summary was developed for the study characteristics and the delivery of the ACASI mode (see Results section).

Methodological Quality Appraisal

The methodological quality of included studies was appraised using the JBI critical appraisal instrument for systematic reviews of prevalence and incidence [32]. There were nine appraisal criteria, which assessed sampling bias, adequacy of sample size, generalizability of study findings, validity of study methods, measurement bias, and the use of appropriate statistical analysis methods. Each criterion was appraised as “yes,” “‘no,” “‘not clear,” or “‘not applicable.” To summarize for reporting purposes, each criterion that was scored as “yes” was assigned a value of 1 and those scored as “no,” “not clear,” or “not applicable” were assigned a value of 0.

Comparison of Survey Modes

To compare sexual behavior outcomes, similar outcome variables were grouped; for example, “buy sex,” “‘sell sex,” “paid for sex,” and “received cash for sex” were grouped together and named “transactional sex.” For each group, the effect sizes that reported on the comparison between the ACASI mode and other survey modes were tabled. Since only a few effect sizes were reported for the comparison between the ACASI mode and the survey modes other than FTFI, only the comparisons between ACASI and FTFI modes were included in the analysis. The effect sizes were grouped based on statistical reports, such as odds ratio (OR), prevalence, predicted percentage, count, mean, and median. When such information was missing, the corresponding authors were contacted for the required information. Because the OR and prevalence values were the most commonly reported, these two types of effect sizes were the focus for this analysis.

When the reviewed papers reported ORs with CIs for the comparison of reporting sexual behaviors between ACASI and FTFI modes, the results were included in the analyses. When the percentages of respondents reporting a sexual behavior or experience and denominators were reported, crude ORs were calculated, with FTFI as the comparative mode. The crude ORs were calculated as the odds of reporting a sexual behavior in the ACASI group divided by the odds of reporting a sexual behavior in the FTFI group. Crude ORs that were greater than 1 were interpreted as follows: the reporting of sexual behaviors was higher with the ACASI mode than with the FTFI mode. For comparisons including biological markers, crude ORs greater than 1 were interpreted as follows: the odds of the discordance between ACASI self-reports and biological marker results were higher than the discordance between FTFI self-reports and biological marker results. For each outcome variable group with eight or more ORs with CIs available, a forest plot was used to visualize the effect sizes reported for each variable group. In addition, the comparison of sexual behavior reports between the ACASI or FTFI mode and biological marker results was charted. No subgroup analysis or meta-analysis was performed due to the heterogeneity of outcome measures and study populations. The heterogeneity for each variable group ranged from 23.8% to 95.7%.

Results

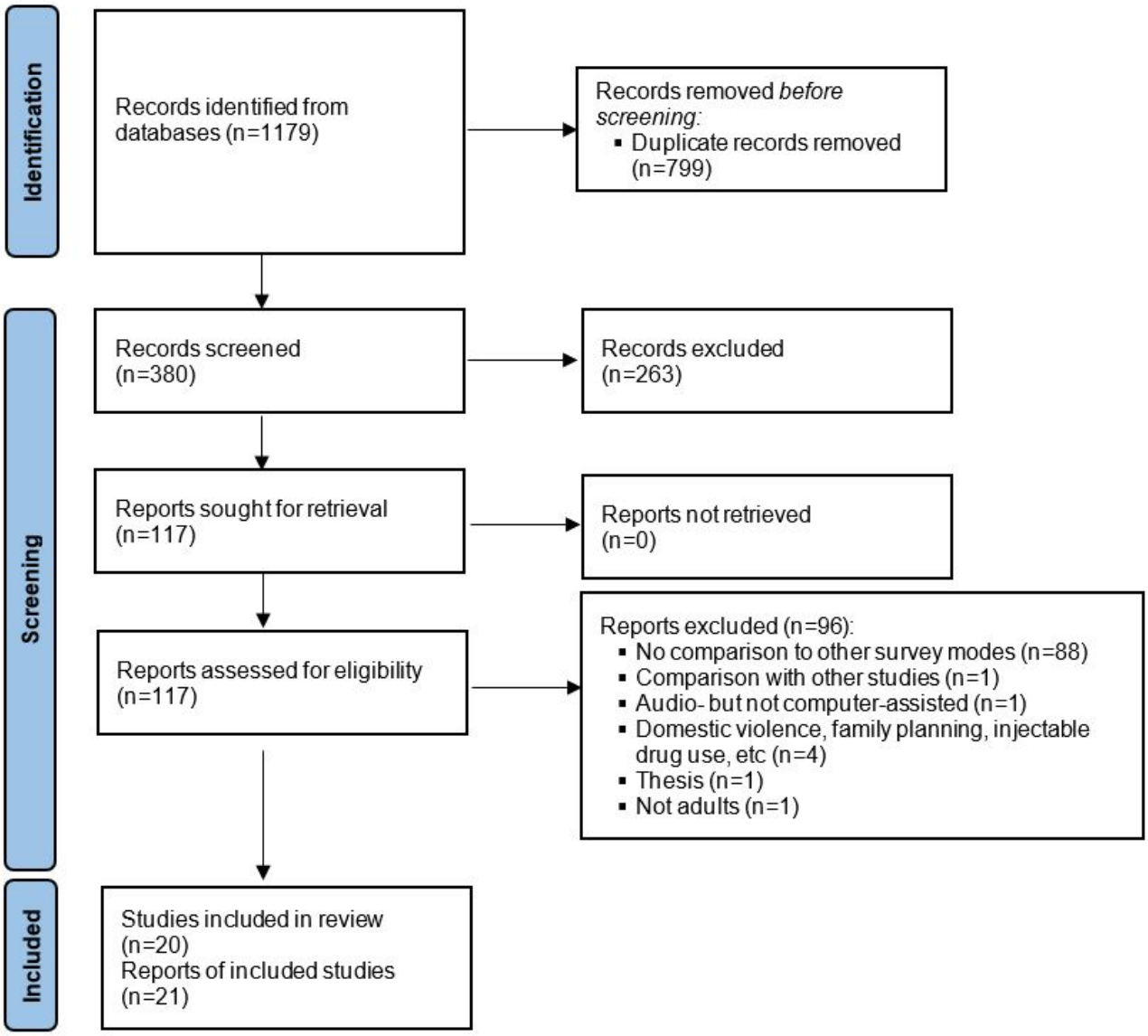
Overview

The search of six databases identified 1179 papers, of which 799 were duplicates (Figure 1). The titles and abstracts of 380 papers were screened, and 263 were removed for reasons such as not testing the ACASI mode, the survey topic not being “sexual behaviors,” the study not originating from Asia or Africa, or participants being younger than 15 years of age. The

XSL•FO
RenderX

full texts of 117 papers were then read and, following this, a further 96 papers were excluded; this left 21 papers. The screening of 650 references cited by the 21 papers, as well as 1049 papers that cited the 21 papers, did not result in the inclusion of any more relevant papers. Among the 21 papers, 2 presented different analysis results of the same study [33,34]. Therefore, a total of 21 papers from 20 different studies were included in this review. The search and screening results are summarized in Figure 1.

Figure 1. 2020 PRISMA flow diagram of the search and screening results of the systematic literature review. The review compared the reporting of sexual behaviors in surveys between those using the ACASI mode and other modes in Asia and sub-Saharan Africa. ACASI: audio computer-assisted self-interview; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Generic Study Characteristics

The generic study characteristics are summarized in Table 1 [33-53]. Among the 20 studies, 65% (n=13) were experimental, quasi-experimental, or randomized trials, while the remaining studies (n=7, 35%) were cross-sectional or longitudinal surveys. More than half of the studies (n=12, 60%) were from sub-Saharan Africa, and the remaining studies (n=8, 40%) were from Asia. The number of participants ranged from 180 to 6530, and all participants were aged 15 years or older. All of the included studies compared the ACASI mode to at least one other survey mode (Table 2). More than half of the studies (n=11, 55%) compared the ACASI mode to only one other survey mode. The FTFI mode was the most compared mode in the reviewed studies (n=17, 85%). In half of the studies (n=10, 50%), the participants used more than one survey mode.

XSL•FO
RenderX

Table 1. Generic study characteristics and quality appraisal from the papers included in the systematic literature review comparing the ACASI[a] mode to other survey modes in sexual behavior surveys in Asia and sub-Saharan Africa.

Author, year	Country	Participants, n	Age (years), range	Females, %	Quality appraisal score[b]
Mensch, 2003 [33]	Kenya	6530	15-21	__[c]	6
Hewett, 2004 [34]	Kenya	709	15-21	100	—
Potdar, 2005 [35]	India	1500	18-22	0	8
Griensven, 2006 [36]	Thailand	1283	15-21	50.1	9
Le, 2006 [37]	Vietnam	2394	15-24	54.3	7
NIMH[d], 2007 [38]	China, India, Peru, Russia and Zimbabwe	445	16-40	48.9	3
Minnis, 2007 [39]	Zimbabwe	655	18-35	100	5
Li, 2007 [40]	China	199	—	45.2	6
Edwards, 2008 [41]	Russia	180	≥18	25	5
Jaya, 2008 [42]	India	1058	15-19	44.9	9
Mensch, 2008 [43]	Malawi	447	15-21	100	7
Minnis, 2009 [44]	Zimbabwe	910	18-49	100	4
van der Elst, 2009 [45]	Kenya	398	—	34.9	3
Mensch, 2011 [46]	South Africa	849	18-40	100	8
Langhaug, 2011 [47]	Zimbabwe	1495	15-23	44.7	8
Le, 2012 [48]	Vietnam	4049	15-49	54.9	7
Gorbach, 2013 [49]	Malawi	585	18-53	100	4
Kelly, 2013 [50]	Malawi	311	16-18	37.3	5
Adebajo, 2014 [51]	Nigeria	1040	≥18	0	3
Kelly, 2014 [52]	Uganda	1020	18-24	100	7
Desmond, 2018 [53]	Malawi	300	16-49	48	6

[a]ACASI: audio computer-assisted self-interview.

[b]The quality appraisal score ranges from 0 to 9.

[c]Not reported.

[d]NIMH: National Institute of Mental Health.

XSL•FO
RenderX

Table 2. Survey modes and outcomes of studies from the papers included in the systematic literature review comparing the ACASI mode to other survey modes in sexual behavior surveys in Asia and sub-Saharan Africa.

Author, year	Modes, n[a]	Modes of survey administration				Modes per participant, n	Outcomes[b]	
		ACASI[c]	FTFI[d]	SAQ[e]	Other modes		Difference	Validation
Mensch, 2003 [33]	3	✓[f]	✓	✓		1	✓	
Hewett, 2004 [34]	2	✓	✓			1	✓	
Potdar, 2005 [35]	3	✓	✓	✓		1	✓	
Griensven, 2006 [36]	4	✓	✓	✓	PASI[g]	1	✓	
Le, 2006 [37]	3	✓	✓	✓		1	✓	
NIMH[h], 2007 [38]	2	✓			CAPI[i]	2 (crossover)	✓	
Minnis, 2007 [39]	2	✓	✓			2 (crossover)	✓	
Li, 2007 [40]	2	✓			CAPI	2 (crossover)	✓	
Edwards, 2008 [41]	2	✓	✓			2	✓	
Jaya, 2008 [42]	3	✓	✓		Interactive[j]	2	✓	
Mensch, 2008 [43]	2	✓	✓			1	✓	
Minnis, 2009 [44]	2	✓	✓		PSA[k]	1	✓	✓
van der Elst, 2009 [45]	2	✓	✓			2	✓	
Mensch, 2011 [46]	2	✓	✓		RSID[l]	1	✓	✓
Langhaug, 2011 [47]	4	✓		✓ (with audio)	ICVI[m]	2	✓	
Le, 2012 [48]	3	✓	✓	✓		1	✓	
Gorbach, 2013 [49]	2	✓	✓			2	✓	
Kelly, 2013 [50]	2	✓	✓			2	✓	
Adebajo, 2014 [51]	2	✓	✓			1	✓	
Kelly, 2014 [52]	3	✓	✓		CAPI and RSID	1	✓	✓
Desmond, 2018 [53]	3	✓	✓ (with visual recall aids)		Daily pictorial diary	3	✓	

[a]Number of modes of survey administration.

[b]"Difference" represents the differences in reporting of sexual behaviors across survey modes; "validation" represents validation using biological markers.

[c]ACASI: audio computer-assisted self-interview.

[d]FTFI: face-to-face interview.

[e]SAQ: self-administered questionnaire.

[f]A checkmark indicates that the respective study included the indicated mode or outcome.

[g]PASI: palmtop-assisted self-interview.

[h]NIMH: National Institute of Mental Health.

[i]CAPI: computer-assisted personal interview.

[j]Interactive interview using audio-visual aids and confidential voting.

[k]PSA: prostate-specific antigen (biological marker).

[l]RSID: rapid stain identification of human semen (biological marker).

[m]ICVI: informal confidential voting interview.

Delivery of the ACASI Mode

Laptops were the most commonly used devices to deliver the ACASI mode (n=9, 45%), while the remaining studies used desktops, handheld computers, and palmtops. In total, 6 (30%) studies used touch-screen devices, and all of these studies were published after 2005. Color- or number-coded keypads were used in 6 (30%) studies. The use of headphones was reported in 12 (60%) studies. The narration was only in a female voice

XSL•FO
RenderX

for 2 (10%) studies, in both genders for 4 (20%) studies, and unreported in the other 14 (70%) studies.

Methodological Quality Appraisal

The sample frame was well defined in the majority of studies (n=17, 85%). Random sampling was used in 15 (75%) studies. The sample size calculation was reported in 8 (40%) studies, and it was adequate in 7 (35%) studies. The study participants and settings were described in detail in the majority of studies (n=18, 90%). The response rate for each survey mode arm was reported in only 7 (35%) studies. The validity of the questionnaire was reported in 9 (45%) studies. The surveys were reliably delivered by trained data collectors in most studies (n=14, 70%). The statistical analysis was appropriate in all the studies. The response rate was 74% and higher in more than half of the studies reviewed (n=13, 65%). On a scale of 0 to 9, the total quality appraisal score for the reviewed studies was 7 to 9 (n=9, 45%), 4 to 6 (n=8, 40%), and 3 (n=3, 15%). The quality appraisal scores are listed in Table 1 [33-53].

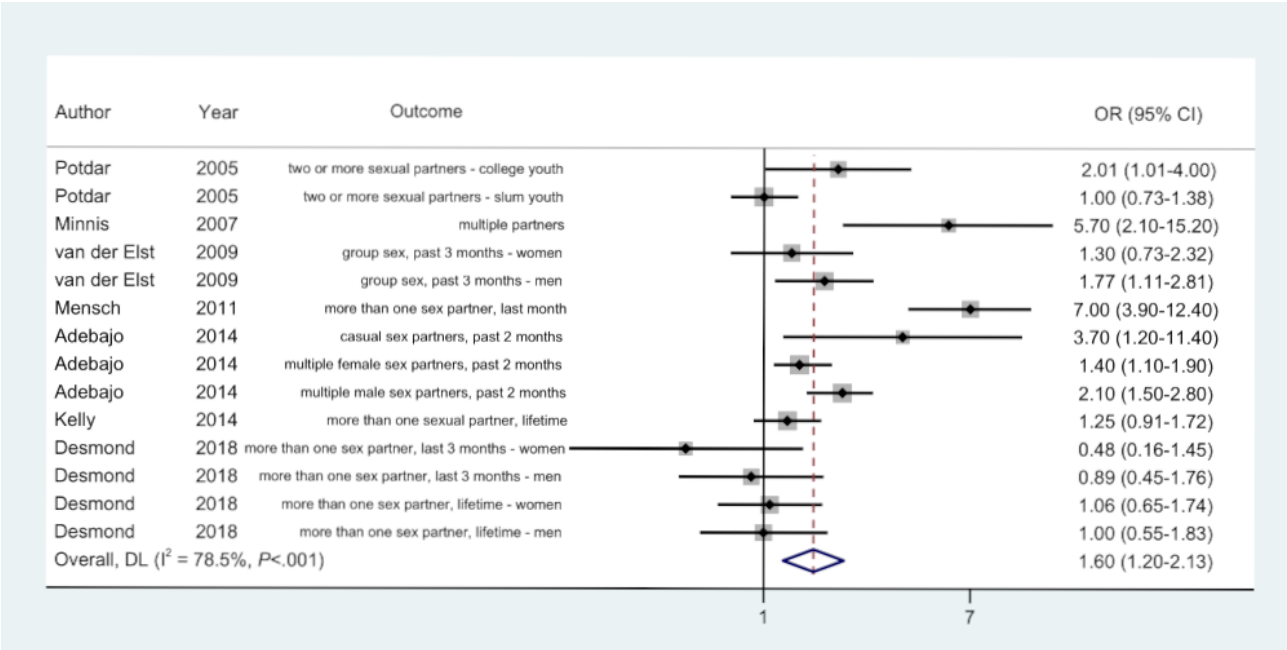
Comparison of Survey Modes

The grouping of similar sexual behavior outcome variables resulted in 23 groups (Table S2 in Multimedia Appendix 1). Among the comparisons between the ACASI mode and other survey modes, the majority were between ACASI and FTFI modes (n=17, 85%). The crude ORs with CIs were available for eight or more comparisons for seven groups of outcome variables, namely, condom use, multiple sexual partners, ever had sex, ever had a partner, transactional sex, anal sex, and forced sex. There were five comparisons between sexual behavior self-reports and biological markers. For these seven groups and the comparison with biological marker results, the effect sizes were visually presented using forest plots (Figures 2-5 and Figures S2-S5 in Multimedia Appendix 1). For the remaining 16 groups of outcome variables, due to the small number of similar outcome measures, the results are summarized in Table S2 in Multimedia Appendix 1, but are not plotted in a figure or included in the narrative report.

Multiple Sexual Partners

A total of 7 out of 20 (35%) studies reported 14 effect sizes for comparing the odds of reporting multiple sexual partners between ACASI and FTFI modes. Participants who used the ACASI mode were more likely to report having multiple sexual partners than those who used the FTFI mode in 10 out of 14 (71%) comparisons, were just as likely in 2 (14%) comparisons, and were less likely in 2 (14%) comparisons (Figure 2). The time frames differed, as follows: the last 1 to 3 months, lifetime, and nonspecified time frame. The wording for "multiple" varied, such as "more than one," "two or more," "multiple," "casual partners," and "group sex." The prevalence of having multiple sexual partners was significantly higher among ACASI participants than FTFI participants in the study that recruited men who had sex with men and men who injected drugs [51]. Similarly, the ACASI participants were 6.4 times (95% CI 3.6-11.2) more likely to report having more than one partner than FTFI participants, and the effect size increased to 7 (95% CI 3.9-12.4) when adjusted for age, marital status, education, ethnicity, and study site (P<.01) [46]. The odds of reporting multiple sexual partners were higher for ACASI participants than FTFI participants in the Zimbabwean study, which reported ORs for the second visit; however, this result might have been affected by the longitudinal crossover design [39]. The results differed across subgroups within the same study. The behavior of having multiple partners was more frequently reported by participants using the ACASI mode than the FTFI mode among college-attending youth in the study in India, whereas there was no such difference for slum youths [35]. Likewise, the difference in the self-reported prevalence of multiple partners between the two modes (ie, ACASI vs FTFI) was significant among male participants but not significant among female participants of the study in Kenya [45].

Figure 2. Random-effects model. Odds ratios (ORs) of reporting "multiple sexual partners" in the ACASI mode as compared to the FTFI mode. ACASI: audio computer-assisted self-interview; DL: DerSimonian-Laird effect size variance estimates; FTFI: face-to-face interview.

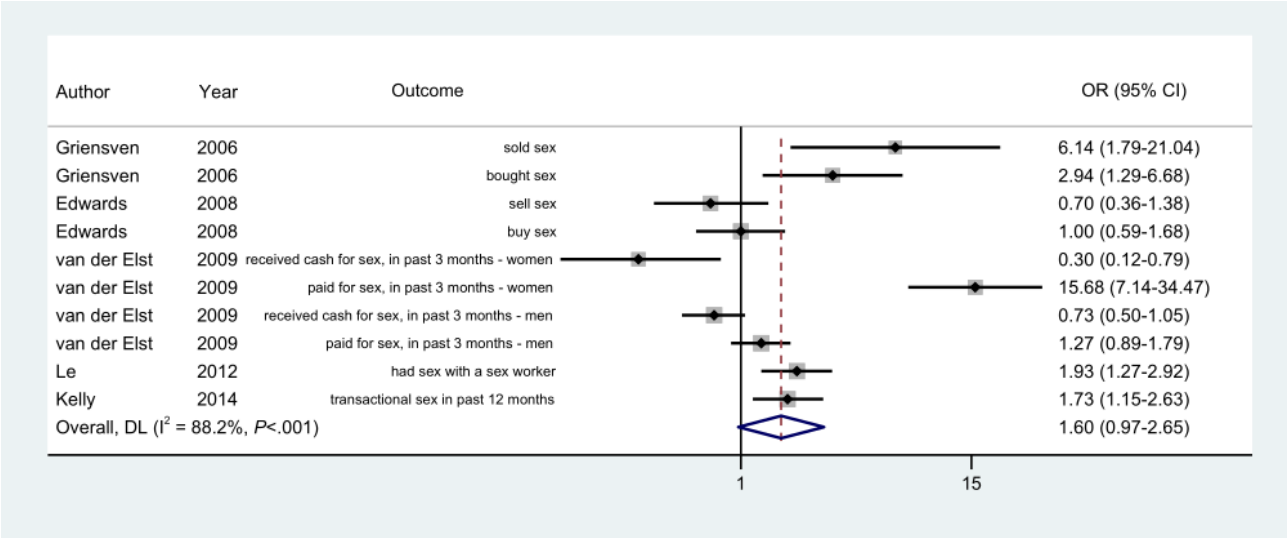


XSL•FO
RenderX

Transactional Sex

The outcome measures related to "transactional sex" were evaluated using ACASI and FTFI modes in 5 out of 20 (25%) studies, and 10 effect sizes were reported (Figure 3). The prevalence of transactional sex was higher in the ACASI arm for 6 (60%) comparisons, it was higher in the FTFI arm for 3 (30%) comparisons, and there was no difference for 1 (10%) comparison. The ACASI participants were more likely to report transactional sex than FTFI participants for 6 (60%) effect sizes, of which 5 (83%) were statistically significant.

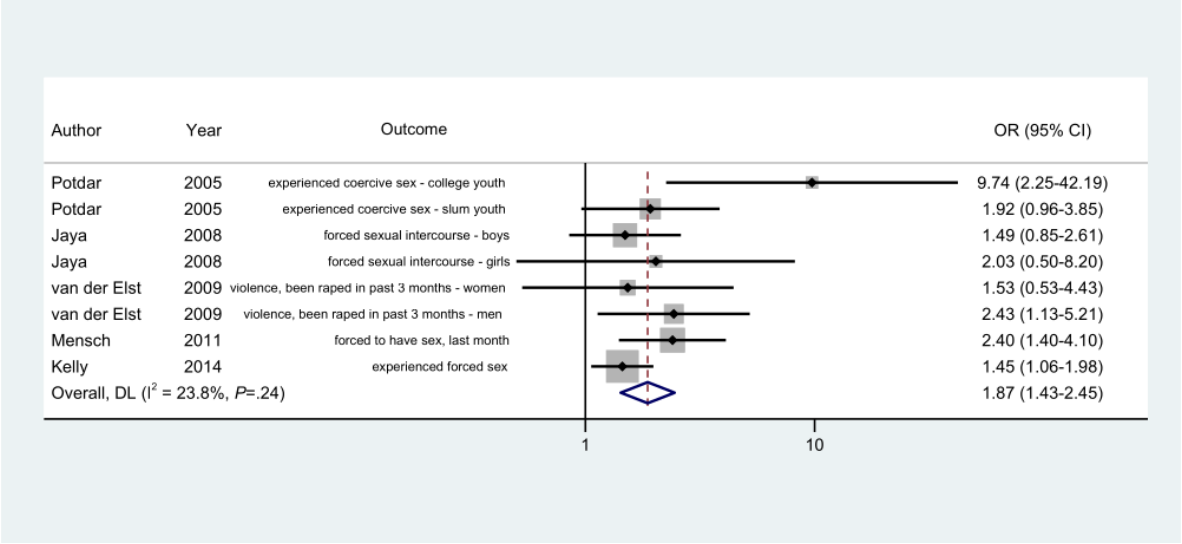
Figure 3. Random-effects model. Odds ratios (ORs) of reporting "transactional sex" in the ACASI mode as compared to the FTFI mode. ACASI: audio computer-assisted self-interview; DL: DerSimonian-Laird effect size variance estimates; FTFI: face-to-face interview.



Forced Sex

The odds of reporting forced sex were higher among ACASI participants than FTFI participants for all 8 results reported by 5 (25%) studies (Figure 4). The wording of outcomes varied, as follows: "forced," "coercive," "violence," and "rape." The higher reports of forced sex among ACASI participants than FTFI participants were significant in the South African and Ugandan studies, in which all participants were women, but both male and female interviewers were involved in the FTFI mode [46,52]. The difference between modes was not statistically significant in the Indian study among unmarried girls and boys aged 15 to 19 years [42]. In the study among sex workers, despite the overlapping CI for women participants, the difference in prevalence was statistically significant (ACASI: 6.6% vs FTFI: 4.4%; n=139; *P*=.40) [45]. The reports of forced sex were statistically significantly higher when the surveys were administered with ACASIs than with FTFIs among college-attending youth but not slum-dwelling youth in India [35].

Figure 4. Random-effects model. Odds ratios (ORs) of reporting "forced sex" in the ACASI mode as compared to the FTFI mode. ACASI: audio computer-assisted self-interview; DL: DerSimonian-Laird effect size variance estimates; FTFI: face-to-face interview.



Condom Use

The odds of reporting condom use were compared between ACASI and FTFI participants in 8 out of 20 (40%) studies, and there were 16 effect sizes. The odds of reporting condom use were lower among ACASI participants than FTFI participants for 7 out of 16 (44%) reports, higher for 7 (44%), and the same for 2 (13%; Figure S2 in Multimedia Appendix 1). Condom use

XSL•FO
RenderX

outcomes had different time frames, such as last sexual encounter, last week, last 3 or 6 months, and nonspecified time frame. There were differences across the studies, such as the study populations, the number of survey modes applied to each participant, and study objectives. All the participants were female in 3 of the 8 (38%) studies [39,49,52]. Populations that were considered high risk were recruited in 2 (25%) studies; these were female and male sex workers in Kenya and patients admitted to the substance abuse treatment hospitals in Russia [41,45]. The participants of the studies in Zimbabwe, Russia, South Africa, and Malawi were administered surveys via both ACASI and FTFI modes [39,41,49,53]. In the study in South Africa, the participants were administered surveys via only one of the two survey modes; however, they had to take the same survey on four occasions [46]. The authors of two microbicide trials noted that participants were likely to overreport condom use due to the adherence counseling they had received [49], and the privacy provided via the ACASI mode alone did not assure disclosure of nonadherence to condom use [46].

Ever Had Sex

The variables related to "ever had sex" were compared between the ACASI and FTFI modes in 5 out of 20 (25%) studies, and 13 effect sizes were reported (Figure S3 in Multimedia Appendix 1). The participants in the ACASI arm reported having had sex more frequently than those in the FTFI arm for 8 (62%) effect sizes, of which only 1 (13%) was statistically significant [35]. Out of 5 studies, 2 (40%) disaggregated the findings by sex, and the higher frequency of reporting ever having had sex among ACASI participants compared to FTFI participants was more prominent among females than males, although the mode difference was not statistically significant [42,50].

Ever Had a Partner

For the outcome variable group "ever had a partner," out of 20 studies, 1 (5%) study in Malawi and 2 (10%) studies in India reported 12 effect sizes (Figure S4 in Multimedia Appendix 1). The odds of reporting ever having had a partner was higher among FTFI participants for 6 (50%) comparisons, higher among ACASI participants for 5 (42%) comparisons, and there was no difference between the two modes for 1 (8%) comparison. In a study in India, the young men recruited from colleges and slums reported same-sex relationships more frequently via the ACASI mode than the FTFI mode, and the mode difference was significant for college youths (OR 7.84, 95% CI 1.78-34.59; *P*=.007) [35].

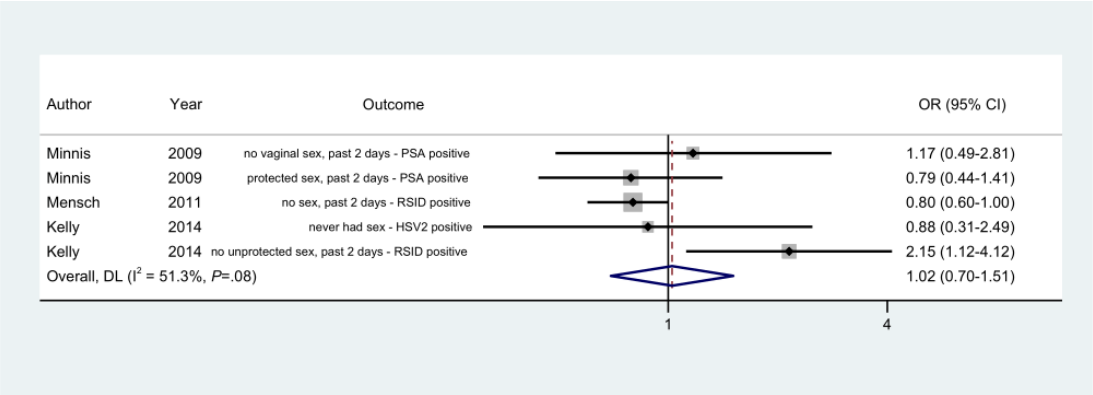
Anal Sex

The survey mode comparison for the outcomes related to "anal sex" was performed in 5 out of 20 (25%) studies, which reported 9 effect sizes (Figure S5 in Multimedia Appendix 1). The reporting of anal sex was more common among ACASI participants than FTFI participants for 4 (44%) effect sizes, and all were statistically significant.

Comparison With Biological Markers

Self-reports of sexual behaviors via the ACASI and FTFI modes were compared to the biological marker results in 3 out of 20 (15%) studies, and 5 outcomes were reported (Figure 5). The biological marker tests used in these studies included herpes simplex virus type 2, prostate-specific antigen, and rapid stain identification of human semen (RSID). The subjective reports of no unprotected sex in the past 2 days and objective findings of the presence of sex were more significantly discordant among ACASI participants than FTFI participants in the study by Kelly et al [52]. The mode difference was marginally significant, with FTFI participants reporting more discordantly with biological marker results than ACASI participants in the study by Mensch et al [46], since ACASI participants were 0.8 times more likely than FTFI participants to report no sex in the past 2 days despite positive RSID results (95% CI 0.6-1.0; *P*=.10).

Figure 5. Random-effects model. Odds ratios (ORs) of the "discordance between self-reports and biological marker results" in the ACASI mode as compared to the FTFI mode. ACASI: audio computer-assisted self-interview; DL: DerSimonian-Laird effect size variance estimates; FTFI: face-to-face interview; HSV2: herpes simplex virus type 2; PSA: prostate-specific antigen; RSID: rapid stain identification of human semen.



Discussion

Principal Findings

A total of 21 papers that reported the results from 20 different studies comparing the reporting of sensitive questions via the ACASI mode and other survey modes were reviewed in this study. The methodological quality of the reviewed papers was moderate to good. The FTFI mode was the most commonly compared mode to ACASI. Among the seven most commonly examined groups of outcome variables, the participants in the ACASI arm were more likely to report having had forced sex, multiple partners, and transactional sex, as well as ever having had sex than the FTFI arm. The odds of reporting were either

XSL•FO
RenderX

higher or lower among ACASI participants compared to FTFI participants for condom use, having a partner, and anal sex. When validated with biological markers for five comparisons, there was no significant difference in three, discordance was higher with ACASI in one, and discordance was higher with FTFI in another.

Comparison With Prior Work

The odds of reporting sexual behaviors, such as forced sex, multiple partners, transactional sex, and having had sex, were higher when the surveys were administered with the ACASI mode than with the FTFI mode. This finding was similar to other reviews that showed higher reporting of sexual behaviors with computer-assisted survey administration [24,27]. A meta-analysis of 48 studies was conducted mostly in the United States and Europe and compared reporting rates of socially undesirable behaviors when using interviewer-delivered and computerized survey modes; the analysis concluded that self-disclosure was higher in the latter, and such differences were more pronounced for highly sensitive behaviors [26]. A meta-analysis of 15 surveys compared reporting rates of HIV risk behaviors in low- and middle-income countries when using FTFIs compared to non-FTFI methods; the analysis reported consistently higher reporting of "forced sex" with non-FTFI methods [25].

Despite higher reporting of sexual behaviors with the ACASI mode compared to the FTFI mode for a few outcome variables, the validation against biological marker results suggested that computerized delivery alone was not always associated with reliable self-reports. This inference was supported by a review of the use of subjective and objective methods in sexual behavior assessment, which concluded that objective assessment was associated with more reliable results [54]. Similarly, a review of survey modes to administer HIV and other STI risk behavior surveys concluded that self-reported prevalence of such risk behaviors was not always high with non-FTFI modes, such as ACASIs, assisted self-administered questionnaires, informal confidential voting interviews, palmtop-assisted self-interviews, polling booth surveys, and tape-recorded interviews [25].

The reviewed studies noted that the self-reported prevalence of forced sex, multiple partners, transactional sex, and having had sex were higher with the ACASI mode than with the FTFI mode; however, the suitability of the use of ACASIs and the self-reports when using the ACASI mode were based on many factors, such as consequences of not involving interviewers, participants' sociodemographic factors, literacy, familiarity with the use of computers or similar devices, participants' conflicts of interest regarding participation in the study, social norms, data collection setting, and study design. These findings were consistent with literature reviews evaluating the impact of inquiry mode on sexual behavior, which reported that many factors, including survey modes, could increase or decrease the reporting of sexual behaviors, and those factors were interactive [22,55]. A few participants highlighted not being able to seek advice from interviewers with the ACASI mode [47]. Similarly, Mensch et al [46] warned that unless a logic check was programmed with the ACASI mode, there would be internally inconsistent responses, which could have been corrected by interviewers via the FTFI mode; for example, when a question about the number of sexual partners was answered with any number greater than zero following a question about ever having had sex being answered as "no," the logic check prompt feature in the ACASI mode would point out the inconsistency and ask the participants to re-enter their response for the previous question.

The heterogeneity of participant characteristics, culture, and social norms within and across the studies reviewed could have confounded the differences across survey modes. Literacy and familiarity with the use of computers and similar devices were participant prerequisites for the ACASI mode. The studies in China, India, and Vietnam concluded that the participants who were familiar with the use of computers were more comfortable taking the ACASI survey [35,40,42,48]. However, in a study in Kenya in 2000, the participants saw the computers as superstitious, especially when it appeared that the computer was talking to them [33]. van der Elst et al [45] remarked that in a population with limited access to health care, study participants could exaggerate risky behaviors such that they would receive benefits in health care through participation in research. The authors of two reviewed studies suggested that sexual behavior reports might have been affected by the cultural conservatism in India and the Confucian culture in Vietnam [35,37].

The data collection setting could also have affected data quality. When data collection took place in a room with around six or seven participants at the same time, the participants felt a lack of privacy, using either ACASI or FTFI modes [35]. In a Russian study, the authors noted that a common problem of data leakage in Russia (eg, selling personal information on the street) could have impaired the study participants' trust in the ACASI mode [41]. In studies with a crossover design, the experience in the first mode could have affected the responses in the subsequent modes [38-40,49]. When participants had to take surveys via both modes on the same day and assumed that their responses would be compared, they might have attempted to keep their responses consistent [41]. Desmond et al [53] suggested that any retrospective survey mode might have an issue with memory recall.

Limitations

There are a few limitations in this study. Since only a limited number of studies from each country or region were identified, the results might not be generalizable to similar populations in corresponding countries or regions. Due to heterogeneity of outcome measures and comparison of the ACASI mode to different survey modes, no meta-analysis or subgroup analysis was performed and, hence, no definitive conclusion was reached. Since this review included only studies published in English, studies published in other languages were missed.

Conclusions

The self-reported prevalence rates of a few sensitive sexual behaviors were higher with the ACASI mode than with the FTFI mode, for example, forced sex, multiple sexual partners, transactional sex, and ever had sex. However, a mix of higher and lower reporting of other sexual behaviors in the ACASI

XSL•FO
RenderX

mode compared to the FTFI mode, as well as discordance between self-reports with the ACASI mode and biological marker results, suggested that an absence of interviewers in the ACASI mode alone did not lower social desirability bias or improve data quality. Moreover, the use of different outcome measures in the reviewed studies hindered the comparison of findings across the studies. In addition to the survey modes, many other factors, such as study design, data collection settings, social norms, and participant characteristics, could have an impact on self-reports of sexual behaviors, and this was not reported in detail by many of the reviewed studies. Therefore, the review findings suggested that the ACASI mode might be a useful survey mode in administering sexual behavior surveys in Asia and sub-Saharan Africa; however, this requires more studies with comparable outcome measures and information about other factors to be taken into consideration in surveying sexual behaviors and to better understand the mode effect of ACASIs in sexual behavior surveys.

Acknowledgments

We would like to thank Corie Gray, who assisted with full-text screening of potentially relevant papers.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 694 KB - jmir_v24i5e37356_app1.docx]

References

1. Key facts HIV. World Health Organization. 2021 Jan 15. URL: https://cdn.who.int/media/docs/default-source/hq-hiv-hepatitis-and-stis-library/key-facts-hiv-2020.pdf [accessed 2022-05-04]
2. Global Hepatitis Report, 2017. Geneva, Switzerland: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/255016/9789241565455-eng.pdf [accessed 2022-01-15]
3. Rowley J, Vander Hoorn S, Korenromp E, Low N, Unemo M, Abu-Raddad LJ, et al. Chlamydia, gonorrhoea, trichomoniasis and syphilis: Global prevalence and incidence estimates, 2016. Bull World Health Organ 2019 Aug 01;97(8):548-562P [FREE Full text] [doi: 10.2471/BLT.18.228486] [Medline: 31384073]
4. Global Progress Report On HIV, Viral Hepatitis And Sexually Transmitted Infections, 2021. Accountability for the Global Health Sector Strategies 2016–2021: Actions for Impact. Geneva, Switzerland: World Health Organization; 2021. URL: https://www.who.int/publications/i/item/9789240027077 [accessed 2022-05-13]
5. Abramson PR. Sexual science: Emerging discipline or oxymoron? J Sex Res 1990 May;27(2):147-165. [doi: 10.1080/00224499009551550]
6. Gagnon JH. Sex research and sexual conduct in the era of AIDS. J Acquir Immune Defic Syndr (1988) 1988;1(6):593-601. [Medline: 3225745]
7. Wellings K, Collumbien M, Slaymaker E, Singh S, Hodges Z, Patel D, et al. Sexual behaviour in context: A global perspective. Lancet 2006 Nov 11;368(9548):1706-1728. [doi: 10.1016/S0140-6736(06)69479-8] [Medline: 17098090]
8. Dean J, Shechter A, Vertkin A, Weiss P, Yaman O, Hodik M, et al. Sexual Health and Overall Wellness (SHOW) survey in men and women in selected European and Middle Eastern countries. J Int Med Res 2013 Apr;41(2):482-492 [FREE Full text] [doi: 10.1177/0300060513476429] [Medline: 23569016]
9. Pinyopornpanish K, Thanamee S, Jiraporncharoen W, Thaikla K, McDonald J, Aramrattana A, et al. Sexual health, risky sexual behavior and condom use among adolescents young adults and older adults in Chiang Mai, Thailand: Findings from a population based survey. BMC Res Notes 2017 Dec 04;10(1):682 [FREE Full text] [doi: 10.1186/s13104-017-3055-1] [Medline: 29202883]
10. Smith AD, Kimani J, Kabuti R, Weatherburn P, Fearon E, Bourne A. HIV burden and correlates of infection among transfeminine people and cisgender men who have sex with men in Nairobi, Kenya: An observational study. Lancet HIV 2021 May;8(5):e274-e283. [doi: 10.1016/S2352-3018(20)30310-6] [Medline: 33631101]
11. Chambliss JT, Evans R, Bolland A, Wingate MS, Bolland JM. Exploring condom use behaviors among African American adolescent boys in the Deep South. Am J Mens Health 2021;15(2):1-10 [FREE Full text] [doi: 10.1177/15579883211009039] [Medline: 33855904]
12. Mutepfa MM, Tapera R. Traditional survey and questionnaire platforms. In: Liamputtong P, editor. Handbook of Research Methods in Health Social Sciences. Singapore: Springer; 2018:1-18.
13. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291. [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
14. Krumpal I. Determinants of social desirability bias in sensitive surveys: A literature review. Qual Quant 2011 Nov 19;47(4):2025-2047. [doi: 10.1007/s11135-011-9640-9]

XSL•FO
RenderX

15. Tourangeau R, Yan T. Sensitive questions in surveys. Psychol Bull 2007 Sep;133(5):859-883. [doi: 10.1037/0033-2909.133.5.859] [Medline: 17723033]
16. Bastien S, Kajula L, Muhwezi W. A review of studies of parent-child communication about sexuality and HIV/AIDS in sub-Saharan Africa. Reprod Health 2011 Sep 24;8:25 [FREE Full text] [doi: 10.1186/1742-4755-8-25] [Medline: 21943095]
17. Griffin S. Literature Review on Sexual and Reproductive Health Rights: Universal Access to Services, Focussing on East and Southern Africa and South Asia. London, UK: Panos London; 2006. URL: https://assets.publishing.service.gov.uk/media/57a08c2940f0b64974001036/LitReview.pdf [accessed 2022-01-15]
18. Ubillos S, Páez D, González JL. Culture and sexual behavior. Psicothema 2000;12:70-82 [FREE Full text]
19. Ussher JM, Perz J, Metusela C, Hawkey AJ, Morrow M, Narchal R, et al. Negotiating discourses of shame, secrecy, and silence: Migrant and refugee women's experiences of sexual embodiment. Arch Sex Behav 2017 Oct;46(7):1901-1921 [FREE Full text] [doi: 10.1007/s10508-016-0898-9] [Medline: 28083724]
20. Fenton K, Johnson A, McManus S, Erens B. Measuring sexual behaviour: Methodological challenges in survey research. Sex Transm Infect 2001 Apr;77(2):84-92 [FREE Full text] [doi: 10.1136/sti.77.2.84] [Medline: 11287683]
21. Catania JA, Binson D, Van Der Straten A, Stone V. Methodological research on sexual behavior in the aids era. Annu Rev Sex Res 1995;6(1):77-125.
22. Catania JA, Gibson DR, Chitwood DD, Coates TJ. Methodological problems in AIDS behavioral research: Influences on measurement error and participation bias in studies of sexual behavior. Psychol Bull 1990 Nov;108(3):339-362. [doi: 10.1037/0033-2909.108.3.339] [Medline: 2270232]
23. Mierzwa S, Souidi S, Savel C. On selecting an appropriate customizable electronic self-report survey research technology. Procedia Eng 2016;159:66-71. [doi: 10.1016/j.proeng.2016.08.065]
24. Langhaug L, Sherr L, Cowan F. How to improve the validity of sexual behaviour reporting: Systematic review of questionnaire delivery modes in developing countries. Trop Med Int Health 2010 Mar;15(3):362-381 [FREE Full text] [doi: 10.1111/j.1365-3156.2009.02464.x] [Medline: 20409291]
25. Phillips AE, Gomez GB, Boily MC, Garnett GP. A systematic review and meta-analysis of quantitative interviewing tools to investigate self-reported HIV and STI associated behaviours in low- and middle-income countries. Int J Epidemiol 2010 Dec;39(6):1541-1555. [doi: 10.1093/ije/dyq114] [Medline: 20630991]
26. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. Behav Res Methods 2015 Dec;47(4):1237-1259. [doi: 10.3758/s13428-014-0533-4] [Medline: 25410404]
27. Simões AMA, Bastos FI. Audio computer-assisted interview: A new technology in the assessment of sexually transmitted diseases, HIV, and drug use [Article in Portuguese]. Cad Saude Publica 2004;20(5):1169-1181 [FREE Full text] [doi: 10.1590/s0102-311x2004000500010] [Medline: 15486659]
28. Hewett PC, Mensch BS, Ribeiro MCSDA, Jones HE, Lippman SA, Montgomery MR, et al. Using sexually transmitted infection biomarkers to validate reporting of sexual behavior within a randomized, experimental evaluation of interviewing methods. Am J Epidemiol 2008 Jul 15;168(2):202-211 [FREE Full text] [doi: 10.1093/aje/kwn113] [Medline: 18525081]
29. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Systematic reviews of prevalence and incidence. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: JBI; 2020.
30. World Bank country and lending groups. The World Bank. URL: https://datahelpdesk.worldbank.org/knowledgebase/articles /906519-world-bank-country-and-lending-groups [accessed 2022-05-13]
31. PRISMA flow diagram. PRISMA. 2021. URL: http://prisma-statement.org/prismastatement/flowdiagram.aspx [accessed 2022-05-13]
32. Munn Z, Moola S, Lisy K, Riitano D, Tufanaru C. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and incidence data. Int J Evid Based Healthc 2015;13(3):147-153. [doi: 10.1097/xeb.0000000000000054]
33. Mensch B, Hewett P, Erulkar A. The reporting of sensitive behavior by adolescents: A methodological experiment in Kenya. Demography 2003 May;40(2):247-268. [doi: 10.1353/dem.2003.0017] [Medline: 12846131]
34. Hewett PC, Mensch BS, Erulkar AS. Consistency in the reporting of sexual behaviour by adolescent girls in Kenya: A comparison of interviewing methods. Sex Transm Infect 2004 Dec;80(Suppl 2):ii43-ii48 [FREE Full text] [doi: 10.1136/sti.2004.013250] [Medline: 15572639]
35. Potdar R, Koenig MA. Does Audio-CASI improve reports of risky behavior? Evidence from a randomized field trial among young urban men in India. Stud Fam Plann 2005 Jun;36(2):107-116. [doi: 10.1111/j.1728-4465.2005.00048.x] [Medline: 15991648]
36. van Griensven F, Naorat S, Kilmarx PH, Jeeyapant S, Manopaiboon C, Chaikummao S, et al. Palmtop-assisted self-interviewing for the collection of sensitive behavioral data: Randomized trial with drug use urine testing. Am J Epidemiol 2006 Feb 01;163(3):271-278. [doi: 10.1093/aje/kwj038] [Medline: 16357109]
37. Le LC, Blum RW, Magnani R, Hewett PC, Do HM. A pilot of audio computer-assisted self-interview for youth reproductive health research in Vietnam. J Adolesc Health 2006 Jun;38(6):740-747. [doi: 10.1016/j.jadohealth.2005.07.008] [Medline: 16730604]

XSL•FO
RenderX

38. NIMH Collaborative HIV/STD Prevention Trial Group. The feasibility of audio computer-assisted self-interviewing in international settings. AIDS 2007 Apr;21 Suppl 2:S49-S58. [doi: 10.1097/01.aids.0000266457.11020.f0] [Medline: 17413264]
39. Minnis AM, Muchini A, Shiboski S, Mwale M, Morrison C, Chipato T, et al. Audio computer-assisted self-interviewing in reproductive health research: Reliability assessment among women in Harare, Zimbabwe. Contraception 2007 Jan;75(1):59-65. [doi: 10.1016/j.contraception.2006.07.002] [Medline: 17161126]
40. Li L, Rotheram-Borus MJ, Batterham P, Lin Y, Wu Z, Detels R, et al. Measuring sexual risk using audio computer-assisted self-interviewing (ACASI) versus computer-assisted personal interview (CAPI) in China. Int J Sex Health 2007 Feb 08;19(1):25-30. [doi: 10.1300/j514v19n01_04]
41. Edwards EM, Cheng DM, Levenson S, Bridden C, Meli S, Egorova VY, et al. Behavioral assessments in Russian addiction treatment inpatients: A comparison of audio computer-assisted self-interviewing and interviewer-administered questionnaires. HIV Clin Trials 2008;9(4):247-253 [FREE Full text] [doi: 10.1310/hct0904-247] [Medline: 18753119]
42. Jaya, Hindin MJ, Ahmed S. Differences in young people's reports of sexual behaviors according to interview methodology: A randomized trial in India. Am J Public Health 2008 Jan;98(1):169-174. [doi: 10.2105/AJPH.2006.099937] [Medline: 18160677]
43. Mensch B, Hewett P, Gregory R, Helleringer S. Sexual behavior and STI/HIV status among adolescents in rural Malawi: An evaluation of the effect of interview mode on reporting. Stud Fam Plann 2008 Dec;39(4):321-334 [FREE Full text] [doi: 10.1111/j.1728-4465.2008.00178.x] [Medline: 19248718]
44. Minnis AM, Steiner MJ, Gallo MF, Warner L, Hobbs MM, van der Straten A, et al. Biomarker validation of reports of recent sexual activity: Results of a randomized controlled study in Zimbabwe. Am J Epidemiol 2009 Oct 01;170(7):918-924 [FREE Full text] [doi: 10.1093/aje/kwp219] [Medline: 19741042]
45. van der Elst EM, Okuku HS, Nakamya P, Muhaari A, Davies A, McClelland RS, et al. Is audio computer-assisted self-interview (ACASI) useful in risk behaviour assessment of female and male sex workers, Mombasa, Kenya? PLoS One 2009;4(5):e5340 [FREE Full text] [doi: 10.1371/journal.pone.0005340] [Medline: 19412535]
46. Mensch BS, Hewett PC, Abbott S, Rankin J, Littlefield S, Ahmed K, et al. Assessing the reporting of adherence and sexual activity in a simulated microbicide trial in South Africa: An interview mode experiment using a placebo gel. AIDS Behav 2011 Feb;15(2):407-421. [doi: 10.1007/s10461-010-9791-z] [Medline: 20886278]
47. Langhaug LF, Cheung YB, Pascoe SJS, Chirawu P, Woelk G, Hayes RJ, et al. How you ask really matters: Randomised comparison of four sexual behaviour questionnaire delivery modes in Zimbabwean youth. Sex Transm Infect 2011 Mar;87(2):165-173. [doi: 10.1136/sti.2009.037374] [Medline: 20943824]
48. Le L, Vu L. Audio computer-assisted self interview compared to traditional interview in an HIV-related behavioral survey in Vietnam. MEDICC Rev 2012;14(4):26-31. [doi: 10.37757/mr2012v14.n4.7]
49. Gorbach PM, Mensch BS, Husnik M, Coly A, Mâsse B, Makanani B, et al. Effect of computer-assisted interviewing on self-reported sexual behavior data in a microbicide clinical trial. AIDS Behav 2013 Feb;17(2):790-800 [FREE Full text] [doi: 10.1007/s10461-012-0302-2] [Medline: 23054034]
50. Kelly CA, Soler-Hampejsek E, Mensch BS, Hewett PC. Social desirability bias in sexual behavior reporting: Evidence from an interview mode experiment in rural Malawi. Int Perspect Sex Reprod Health 2013 Mar;39(1):14-21 [FREE Full text] [doi: 10.1363/3901413] [Medline: 23584464]
51. Adebajo S, Obianwu O, Eluwa G, Vu L, Oginni A, Tun W, et al. Comparison of audio computer assisted self-interview and face-to-face interview methods in eliciting HIV-related risks among men who have sex with men and men who inject drugs in Nigeria. PLoS One 2014;9(1):e81981 [FREE Full text] [doi: 10.1371/journal.pone.0081981] [Medline: 24416134]
52. Kelly CA, Hewett PC, Mensch BS, Rankin JC, Nsobya SL, Kalibala S, et al. Using biomarkers to assess the validity of sexual behavior reporting across interview modes among young women in Kampala, Uganda. Stud Fam Plann 2014 Mar;45(1):43-58 [FREE Full text] [doi: 10.1111/j.1728-4465.2014.00375.x] [Medline: 24615574]
53. Desmond N, Nagelkerke N, Lora W, Chipeta E, Sambo M, Kumwenda M, et al. Measuring sexual behaviour in Malawi: A triangulation of three data collection instruments. BMC Public Health 2018 Jun 28;18(1):807 [FREE Full text] [doi: 10.1186/s12889-018-5717-x] [Medline: 29954360]
54. Mirzaei M, Ahmadi K, Saadat S, Ramezani M. Instruments of high risk sexual behavior assessment: A systematic review. Mater Sociomed 2016 Feb;28(1):46-50 [FREE Full text] [doi: 10.5455/msm.2016.28.46-50] [Medline: 27047267]
55. McCallum EB, Peterson ZD. Investigating the impact of inquiry mode on self-reported sexual behavior: Theoretical considerations and review of the literature. J Sex Res 2012;49(2-3):212-226. [doi: 10.1080/00224499.2012.658923] [Medline: 22380589]

Abbreviations

ACASI: audio computer-assisted self-interview
BBV: blood-borne virus
FTFI: face-to-face interview
JBI: Joanna Briggs Institute

XSL•FO
RenderX

OR: odds ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RSID: rapid stain identification of human semen
STD: sexually transmitted disease
STI: sexually transmitted infection
WHO: World Health Organization

Edited by R Kukafka; submitted 17.02.22; peer-reviewed by J Gauly, R Bajpai, M Tran; comments to author 07.04.22; revised version received 20.04.22; accepted 21.04.22; published 31.05.22.

Please cite as:
Phoo NNN, Lobo R, Vujcich D, Reid A
Comparison of the ACASI Mode to Other Survey Modes in Sexual Behavior Surveys in Asia and Sub-Saharan Africa: Systematic Literature Review
J Med Internet Res 2022;24(5):e37356
URL: https://www.jmir.org/2022/5/e37356
doi:10.2196/37356
PMID:35639465

©Nang Nge Nge Phoo, Roanna Lobo, Daniel Vujcich, Alison Reid. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2022. 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.

XSL•FO
RenderX

Commentary

Through the Narrative Looking Glass: Commentary on "Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review"

Charlene Weir[1], PhD

Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Charlene Weir, PhD
Department of Biomedical Informatics
University of Utah
421 Wakara Way, Ste 140
Salt Lake City, UT, 84108-3514
United States
Phone: 1 801 541 9462
Email: charlene.weir@utah.edu

Related Article:

Comment on: http://www.jmir.org/2022/5/e30405/

Abstract

The authors of "Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review" have effectively brought to our attention the failure of the electronic health record (EHR) to represent the human context. Because mental health or behavioral disorders (and functional status in general) emerge from an interaction between the individual's characteristics and the social context, it is essentially a failure to represent the human context. The assessment and treatment of these disorders must reflect how the person lives, their degree of social connectedness, their personal motivation, and their cultural background. This type of information is best communicated both through narrative and in collaboration with other providers and the patient—largely because human social memory is organized around situation models and natural episodes. Neither functionality is currently available in most EHRs. Narrative communication is effective for several reasons: (1) it supports the communication of goals between providers; (2) it allows the author to express their belief in others' perspectives (theory of mind), for example, those who will be reading these notes; and (3) it supports the incorporation of the patient's personal perspective. The failure of the EHR to support mental health information data and information practices is, therefore, essentially a failure to support the basic communication functions necessary for the narrative. The authors have rightly noted the problems of the EHR in this domain, but perhaps they did not completely link the problems to the lack of functionality to support narrative communication. Suggestions for adding design elements are discussed.

(J Med Internet Res 2022;24(5):e38513) doi:10.2196/38513

KEYWORDS

electronic health records; psychiatry; mental health; electronic medical records; health informatics; mental illness; scoping review; clinical decision support

Introduction

Through their scoping review of mental health data and the electronic health record (EHR), the authors of "Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review" have brought to center stage the failure of the EHR to represent the human context [1]. Mental health or behavioral disorders (and functional status in general) emerge from an *interaction* between the individual's characteristics and the social context. As a result, the assessment and treatment of these disorders must reflect how the person lives, their degree of social connectedness, their personal motivation, and their cultural background—in other words: the human context. This failure of the EHR to support both information data (eg, missing or "fuzzy" data) and information practices (processes) for mental health information is a feature, not a bug. Specifically, EHRs have systematically avoided

XSL•FO
RenderX

supporting text data—partially because electronic text is seen as hard to use [2] and due to the belief that structured data is more accurate. However, it is not just that providers *prefer* to tell the patient's story in narrative rather than structured data forms [3] or that mental health data is "soft" data, it is that it is much too difficult to get a sense or gist of the patient's situation through structured data and is much less cognitively efficient. In other words, accuracy is in the eye of the beholder. Some studies have noted the narrative is more accurate for mental health data, even if different text is used as descriptors [4]. One reason for the power of the narrative is that memory is organized around situation models and episodic mental representations, which are best communicated in story form [5]. Humans can grasp a situation much more rapidly through a story than through a list of facts [5]. Putting together "pieces" of data to get a gist of the patient's situation is significant work, whereas distilled information has better comprehension and is associated with better decision-making [6]. Narrative communication is effective for several reasons: (1) it supports the communication of goals between providers; (2) it allows for the author to express their belief in others' perspectives (theory of mind), for example, those who will be reading these notes; and (3) it supports the incorporation of the patient's personal perspective.

Goals

Documenting and tracking clinical goals is at the heart of care processes and communication in general [7]. The clinical goals for mental health patients almost always involve some aspect of context (which, in turn, requires specific descriptions of that context). The question "is the therapeutic treatment working?" requires data about the patient's work situation, personal relationships, or the patient's motivation [8].

Communication

Communicating mental health information to other providers is complex because people of many different roles care for these patients compared to patients with other disorders. Documentation must then be "tailored" to the audience and to the perspectives of differing roles (theory of mind), which requires significant amounts of working memory [9].

Patient Preferences

The patient's preferences are often idiosyncratic, embedded in the social context, and specific to location and time. The EHR is a limited representation of patient preference data. A story about the patient's wishes is generally the most effective way of communicating preferences and planning care [10].

Conclusion

The failure of the EHR to support mental health information data and information practices is, therefore, really a failure to support the basic communication functions necessary for the narrative. The authors have rightly noted the problems of the EHR in this domain, but perhaps they did not completely link the problems to the lack of functionality to support narrative communication. Links to the clinical goals of other clinicians, a specific location for the patient's story, temporal links to clinical episodes, and the ability to annotate the clinical notes of others in order to understand one's impressions would help communicate the patient's story. Improving the use of natural language processing and building ontologies of context would also help. Additionally, addressing these functions would also address several of the issues raised in the review, specifically, missing data, sensitive data, and collaborative decision-making information. Future work in the arena of EHRs could create tools and spaces for narrative, patient preferences, collaborative discourse, and shared collaborative documentation [11].

Conflicts of Interest

None declared.

References

1. Kariotis T, Prictor M, Chang S, Gray K. Impact of Electronic Health Records on Information Practices in Mental Health Context: Scoping Review. J Med Internet Res 2022;24(5):e30405. [doi: 10.2196/30405]
2. Embi PJ, Weir C, Efthimiadis EN, Thielke SM, Hedeen AN, Hammond KW. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc 2013;20(4):718-726 [FREE Full text] [doi: 10.1136/amiajnl-2012-000946] [Medline: 23355462]
3. Rizvi RF, Harder KA, Hultman GM, Adam TJ, Kim M, Pakhomov SVS, et al. A comparative observational study of inpatient clinical note-entry and reading/retrieval styles adopted by physicians. Int J Med Inform 2016 Jun;90:1-11 [FREE Full text] [doi: 10.1016/j.ijmedinf.2016.02.011] [Medline: 27103191]
4. Spuhl J, Doing-Harris K, Nelson S, Estrada N, Del Fiol G, Weir C. Concordance of Electronic Health Record (EHR) Data Describing Delirium at a VA Hospital. AMIA Annu Symp Proc 2014;2014:1066-1071 [FREE Full text] [Medline: 25954416]
5. Wyer Jr RS. Social Comprehension and Judgment: The Role of Situation Models, Narratives, and Implicit Theories. Mahwah, New Jersey: Lawrence Erlbaum Associates; Jan 11, 2013. [doi: 10.4324/9781410609007]
6. Blalock SJ, Reyna VF. Using fuzzy-trace theory to understand and improve health judgments, decisions, and behaviors: A literature review. Health Psychol 2016 Aug;35(8):781-792 [FREE Full text] [doi: 10.1037/hea0000384] [Medline: 27505197]
7. Clarke HH. Using Language. Cambridge, UK: Cambridge University Press; 1996.

XSL•FO
RenderX

8. Cimino J. Putting the "why" in "EHR": capturing and coding clinical cognition. J Am Med Inform Assoc 2019 Nov 01;26(11):1379-1384 [FREE Full text] [doi: 10.1093/jamia/ocz125] [Medline: 31407781]
9. Senay I, Keysar B. Keeping Track of Speaker's Perspective: The Role of Social Identity. Discourse Processes 2009 Sep 14;46(5):401-425. [doi: 10.1080/01638530902959596]
10. Bekker HL, Winterbottom AE, Butow P, Dillard AJ, Feldman-Stewart D, Fowler FJ, et al. Do personal stories make patient decision aids more effective? A critical review of theory and evidence. BMC Med Inform Decis Mak 2013 Nov 29;13(S2). [doi: 10.1186/1472-6947-13-s2-s9]
11. Stanhope V, Ingoglia C, Schmelter B, Marcus SC. Impact of person-centered planning and collaborative documentation on treatment adherence. Psychiatr Serv 2013 Jan;64(1):76-79. [doi: 10.1176/appi.ps.201100489] [Medline: 23280459]

Abbreviations

EHR: electronic health record

Edited by T Leung; submitted 05.04.22; this is a non–peer-reviewed article;accepted 05.04.22; published 04.05.22.

Please cite as:
Weir C
Through the Narrative Looking Glass: Commentary on "Impact of Electronic Health Records on Information Practices in Mental Health Contexts: Scoping Review"
J Med Internet Res 2022;24(5):e38513
URL: https://www.jmir.org/2022/5/e38513
doi:10.2196/38513
PMID:35507399

©Charlene Weir. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 04.05.2022. 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.

XSL•FO
RenderX

Viewpoint

Improving Research Patient Data Repositories From a Health Data Industry Viewpoint

Chunlei Tang[1*], PhD; Jing Ma[2*], MD, PhD; Li Zhou[1], MD, PhD; Joseph Plasek[1], PhD; Yuqing He[3], MSE; Yun Xiong[4], PhD; Yangyong Zhu[4], PhD; Yajun Huang[3], PhD; David Bates[1], MD

[1]Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[2]Sichuan Academy of Medical Sciences and Sichuan Provincial People's Hospital, Sichuan, China

[3]School of Economics, Fudan University, Shanghai, China

[4]School of Computer Science, Fudan University, Shanghai, China

[*]these authors contributed equally

Corresponding Author:
Chunlei Tang, PhD
Brigham and Women's Hospital
Harvard Medical School
1620 Tremont Street, BS-3
Brigham and Women's Hospital
Boston, MA, 02115
United States
Phone: 1 857 366 7211
Email: towne.tang@gmail.com

Abstract

Organizational, administrative, and educational challenges in establishing and sustaining biomedical data science infrastructures lead to the inefficient use of Research Patient Data Repositories (RPDRs). The challenges, including but not limited to deployment, sustainability, cost optimization, collaboration, governance, security, rapid response, reliability, stability, scalability, and convenience, restrict each other and may not be naturally alleviated through traditional hardware upgrades or protocol enhancements. This article attempts to borrow data science thinking and practices in the business realm, which we call the data industry viewpoint, to improve RPDRs.

(J Med Internet Res 2022;24(5):e32845) doi:10.2196/32845

KEYWORDS

data science; big data; data mining; data warehousing; information storage and retrieval

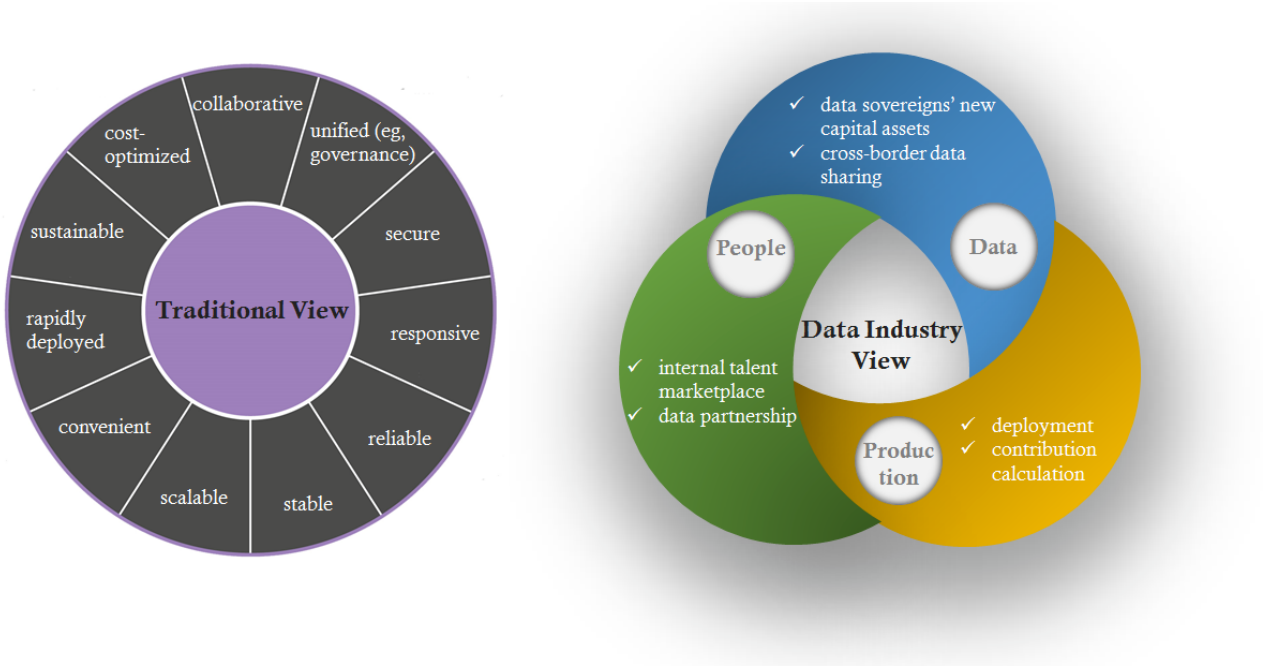
Introduction

Research Patient Data Repositories (RPDRs, eg, Integrating Biology & the Bedside [i2b2]) and their rapid organic evolution are critical to linking disparate and high-dimensional patient data for a wide range of applications in research. One goal for RPDRs' evolution in clinical and translational science is to subsume biomedical data science infrastructures and infrastructural health data science [1,2], such as rapid pharmacovigilance [3] and the delivery of real-world evidence at the point of care to actualize the learning health care system [4]. The path to achieving this goal may be tortuous since problems may not emerge until fundamental issues are resolved. Biomedical data science aims to use data technology of any kind to advance medical society as a transdisciplinary ecosystem [4,5] by unifying different disciplines beyond their traditional boundaries to address a common problem. The complexity of the data science ecosystem increases the difficulty of improving RPDRs. Improving RPDRs, therefore, requires a wide variety of new functions and capabilities in the administrative, organizational, and educational areas, including data integration, management, education, support, tooling, governance, optimization, and alignment across missions [3,6].

The effort to establish and sustain biomedical data science infrastructures would benefit if it borrowed thinking and best practices from the data industry [7]. Data industry thinking includes perspectives on data-driven research, innovation, industrialization, and opportunities. We hypothesize that data industry thinking may reshape prevailing views of how people interact with data value and data production in the context of RPDRs (Figure 1).

XSL•FO
RenderX

Figure 1. Comparison between traditional and data industry viewpoints of Research Patient Data Repositories.



Data Production: Deployment Challenges and Contribution Calculations

Data production involves the generation, storage, and curation of data from data-centric human (social, economic, and scientific) activities. Intuitively speaking, it is the process of combining various analyzable data inputs for consumption. The consumption process starts with incoming raw materials used for the preparation of semifinished (eg, pretrained word embeddings) and finished data products (eg, a service). The raw data and data products are "nonrivalrous" in nature, meaning they can be used by multiple users at once without depletion of the resource. Data products can act as reusable resources [8], assets [9], or capital [10] to accelerate research.

When considering data production in RPDRs, some previously unseen problems may arise, such as deployment. Campion Jr et al [11] reported that deployment challenges are widespread in the existing RPDRs: "a number of tools commonly but not uniformly implemented"; for example, i2b2 enables investigators to obtain deidentified patient counts without SQL programming [12]. Many incorrectly think of deploying a data science or analytical model as the last stage of the process. Starting with the algorithm first, and only at the end of the project thinking about how to insert it into the process, is where many deployments fail [13]. Scientists can readily interact with RPDRs to access the underlying electronic health record (EHR) data. RPDRs should additionally provide a solution for fully and successfully implementing analytical and artificial intelligence models from experimentation to production. The first tools to consider to mitigate deployment challenges are tools for handling structured and unstructured EHR data, such as exploratory analysis and data self-governance tools. Exploratory data analysis is an important data industry best practice step focused on gaining insights from raw data prior to training learning models. Exploratory analysis tools that go beyond basic initial data analysis tasks (like SQL programming, ie, sort, filter, aggregate, correlate, group, derive attributes) are essential for handling tasks that previously were manual, heuristic-based, or simply impossible [14]. The transformation of unstructured clinical notes which contain summaries (eg, history of present illness) that describe and illustrate the longitudinal course of specific clinical events or situations experienced by patients into an appropriate data representation (eg, annotated corpus of pretrained word embeddings or a hierarchical representation with multiple levels of granularity) can offer RPDRs enhanced machine intelligence for downstream analysis and reduce duplicated preprocessing efforts to make this data computable [15]. Data self-governance models like Databox [16] can support data sharing that meets study eligibility criteria documented in RPDRs. These default tools can be customized as digital "errand runners" [17] to replace deeply occupational tasks that are tedious, time-consuming, and not artistic.

Data product sharing should be encouraged by the data sovereigns of RPDRs [18], including cross-border data flows. Multilevel data products, such as models, code, intermediate results, annotated training corpora, enclaves, experimental findings, presentations, preprints, and retrieved literature citations can be found throughout the entire life cycle of medical research and are helpful for accelerating complementary efforts. We recommend transplanting contribution margin–based pricing from the data industry to RPDRs to facilitate data sharing. These contributions include but are not limited to reuse frequency, shareable integrity, quantity versus speed in question and answer responses, and compliance practices. Contribution calculations can support employee engagement in the RPDR community and serve as an accelerator for scientific discovery.

XSL•FO
RenderX

People: Internal Talent Marketplace and Data Partnerships

We suggest that RPDR processes and structures be optimized based on the organizational structure, how stakeholder power is exercised, how stakeholders communicate their needs, how decisions are made, and how decision-makers are held accountable. Data production relies on the efforts of a community of interdisciplinary users, including data scientists, enterprise information technology personnel, clinicians, researchers, informaticists, data engineers, data analysts, annotators, and other data product enhancers. The data partnerships' teams rely on an organization's brand to undertake and complete data production. These teams can freely use RPDR data within organizations, and products or services carried out by these teams will be shared within the company. When the velocity of data partnerships in a market exceeds that of an organization, inefficiencies will cause the organization to lose competitive advantages. As markets evolve, an organization will inevitably choose to focus on cost (ie, replacing human labor with machines) or evolve their organizational structure. Flattening the organizational hierarchy so that people can work together “more equally” will lead to increased efficiencies from equitable data partnerships and the rise of the internal talent marketplace. As an upgraded version of a “principal investigator,” a data partnership might not just rely on grants but also on contributions. In essence, the organization has evolved into a market with relatively small competition. Crowdsourcing within an organization is an alternative for these teams to achieve their goals and with it, the rise of the internal talent marketplace is achieved. The internal talent marketplace takes advantage of the increased flexibility of the gig economy and marketplace-based platforms without requiring changes to employment categories. It matches internal employees and, in some cases, a pool of contingent workers to short-term projects and work. Thus, under ideal next-generation RPDRs, these trends among employees can result in collaborative translational medicine by maintaining an innovation ecosystem through teamwork, trust, reliability, and collaboration.

Conclusions

Best practices in RPDRs tend to focus on core infrastructural and methodological needs, such as machine-readable standards, data access platforms, search and discoverability, claim validation, and insight generation [19]; we argue that the complementary data industry viewpoint is relevant and apposite. From this point of view, RPDRs must consider production deployment and contribution calculations, the establishment of internal talent marketplaces and data partnerships, as well as data sovereigns’ new capital assets and cross-border data sharing, as they reveal issues that are not typically addressed. Only with innovative deployed tools, the wide availability and use of diverse data products, and achievable foresight will the future of ideal next-generation RPDRs be truly accessible.

Authors' Contributions

All authors provided substantial contributions to paper conception and edits and approved the final version of the manuscript.

Conflicts of Interest

JP reports receiving personal fees from Summary Medical Inc and DispatchHealth and equity from Summary Medical Inc outside the submitted work. DB reports receiving grants and personal fees from EarlySense, personal fees from CDI Negev, equity from Valera Health, equity from CLEW Medical, equity from MDClone, personal fees and equity from AESOP, personal fees and equity from FeelBetter, and grants from IBM Watson Health, outside the submitted work.

References

1. Data infrastructure. National Institutes of Health Office of Data Science Strategy. URL: https://datascience.nih.gov/data-infrastructure [accessed 2021-06-11]
2. Meng XL. Building data science infrastructures and infrastructural data science. Harvard Data Science Review. 2021 Apr 30. URL: https://hdsr.mitpress.mit.edu/pub/kdqoo5ax [accessed 2021-06-11]
3. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. Health Aff (Millwood) 2014 Jul;33(7):1229-1235. [doi: 10.1377/hlthaff.2014.0099] [Medline: 25006150]
4. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. J Am Med Inform Assoc 2012;19(2):181-185 [FREE Full text] [doi: 10.1136/amiajnl-2011-000492] [Medline: 22081225]
5. Newman MEJ. The structure of scientific collaboration networks. Proc Natl Acad Sci U S A 2001 Jan 16;98(2):404-409 [FREE Full text] [doi: 10.1073/pnas.98.2.404] [Medline: 11149952]
6. Jacobs JA. Why the disciplines still matter. The Chronicle of Higher Education. 2014 May 27. URL: https://www.chronicle.com/article/Why-the-Disciplines-Still/146777 [accessed 2021-06-11]
7. Tang C. The Data Industry: The Business and Economics of Information and Big Data, 1st Edition. Hoboken, NJ: John Wiley & Sons; 2016.
8. The world’s most valuable resource is no longer oil, but data. The Economist. 2017 May 06. URL: https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource [accessed 2021-06-11]

XSL•FO
RenderX

9. Fisher T. The Data Asset: How Smart Companies Govern Their Data for Business Success, 1st Edition. Hoboken, NJ: John Wiley & Sons; 2009.
10. Tang C. Data Capital: How Data is Reinventing Capital for Globalization, 1st Edition. Berlin/Heidelberg, Germany: Springer; 2021.
11. Campion J, Craven CK, Dorr DA, Knosp BM. Understanding enterprise data warehouses to support clinical and translational research. J Am Med Inform Assoc 2020 Jul 01;27(9):1352-1358 [FREE Full text] [doi: 10.1093/jamia/ocaa089] [Medline: 32679585]
12. 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;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
13. Davenport T, Malone K. Deployment as a critical business data science discipline. Harvard Data Science Review. 2021 Feb 10. URL: https://doi.org/10.1162/99608f92.90814c32 [accessed 2021-06-11]
14. Ghosh A, Nashaat M, Miller J, Quader S, Marston C. A comprehensive review of tools for exploratory analysis of tabular industrial datasets. Vis Inform 2018 Dec;2(4):235-253. [doi: 10.1016/j.visinf.2018.12.004]
15. Weng WH, Szolovits P. Representation learning for electronic health records. arXiv. Preprint posted online on September 19, 2019 2019 (forthcoming) [FREE Full text]
16. Zhu Y, Xiong Y, Liao Z. Self-governing openness of data. Big Data Res 2018;4(2):3-14.
17. Mayer-Schönberger V, Ramge T. Reinventing Capitalism in the Age of Big Data. New York, NY: Basic Books; 2018.
18. Tang C, Plasek JM, Zhu Y, Huang Y. Data sovereigns for the world economy. Humanit Soc Sci Commun 2020 Dec 16;7(184). [doi: 10.1057/s41599-020-00664-y]
19. Yarkoni T, Eckles D, Heathers JAJ, Levenstein MC, Smaldino PE, Lane J. Enhancing and accelerating social science via automation: challenges and ppportunities. Harvard Data Science Review. 2021. URL: https://doi.org/10.1162/99608f92.df2262f5 [accessed 2021-06-11]

Abbreviations

EHR: electronic health record
RPDR: Research Patient Data Repositories

Edited by A Mavragani; submitted 11.08.21; peer-reviewed by P Zhao, W Dalton; comments to author 03.11.21; revised version received 12.01.22; accepted 07.02.22; published 11.05.22.

Please cite as:
Tang C, Ma J, Zhou L, Plasek J, He Y, Xiong Y, Zhu Y, Huang Y, Bates D
Improving Research Patient Data Repositories From a Health Data Industry Viewpoint
J Med Internet Res 2022;24(5):e32845
URL: https://www.jmir.org/2022/5/e32845
doi:10.2196/32845
PMID:35544299

©Chunlei Tang, Jing Ma, Li Zhou, Joseph Plasek, Yuqing He, Yun Xiong, Yangyong Zhu, Yajun Huang, David Bates. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.05.2022. 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.

XSL•FO
RenderX

Viewpoint

Measuring Health-Related Quality of Life With Multimodal Data: Viewpoint

Ieuan Clay[1], PhD; Francesca Cormack[2], PhD; Szymon Fedor[3], PhD; Luca Foschini[4], PhD; Giovanni Gentile[5], PhD; Chris van Hoof[6], PhD; Priya Kumar[7], BA; Florian Lipsmeier[8], PhD; Akane Sano[9], PhD; Benjamin Smarr[10], PhD; Benjamin Vandendriessche[11], PhD; Valeria De Luca[12], PhD

[1]Digital Medicine Society, Boston, MA, United States
[2]Cambridge Cognition, Cambridge, United Kingdom
[3]MIT Media Lab, Cambridge, MA, United States
[4]Evidation Health, San Mateo, CA, United States
[5]University of Padua, Padua, Italy
[6]imec, Leuven, Belgium
[7]Rune Labs, San Fransisco, CA, United States
[8]pRED, Hoffmann-La Roche, Basel, Switzerland
[9]Department of Electrical and Computer Engineering, Rice University, Houston, TX, United States
[10]Department of Bioengineering and Halicioglu Data Science Institute, University of California, San Diego, San Diego, CA, United States
[11]Byteflies, Antwerp, Belgium
[12]Novartis Institutes for Biomedical Research, Basel, Switzerland

Corresponding Author:
Valeria De Luca, PhD
Novartis Institutes for Biomedical Research
Fabrikstrasse 2
Basel, 4056
Switzerland
Phone: 41 79 540 3272
Email: valeria.de_luca@novartis.com

Abstract

The ability to objectively measure aspects of performance and behavior is a fundamental pillar of digital health, enabling digital wellness products, decentralized trial concepts, evidence generation, digital therapeutics, and more. Emerging multimodal technologies capable of measuring several modalities simultaneously and efforts to integrate inputs across several sources are further expanding the limits of what digital measures can assess. Experts from the field of digital health were convened as part of a multi-stakeholder workshop to examine the progress of multimodal digital measures in two key areas: detection of disease and the measurement of meaningful aspects of health relevant to the quality of life. Here we present a meeting report, summarizing key discussion points, relevant literature, and finally a vision for the immediate future, including how multimodal measures can provide value to stakeholders across drug development and care delivery, as well as three key areas where headway will need to be made if we are to continue to build on the encouraging progress so far: collaboration and data sharing, removal of barriers to data integration, and alignment around robust modular evaluation of new measurement capabilities.

(J Med Internet Res 2022;24(5):e35951) doi:10.2196/35951

KEYWORDS

digital measures; quality of life; machine learning; digital health; digital product; digital wellness; digital therapeutics; digital therapy; multimodal technology; drug development; care delivery; data integration

Introduction

The field of digital health has become a multibillion dollar market, powering a paradigm shift by enabling the continuous capture of multimodal data including activity, sleep, vital signs, and contextual information. Novel machine learning applications are pioneering the conversion of these multimodal data into measures for health-related quality of life (QOL)–relevant

XSL•FO
RenderX

symptoms like fatigue [1], stress [2], and depression [3,4]. These insights have the potential to enable improved care delivery [5] and a deeper understanding of patients' lived experiences and better, more personalized medicines. However, important barriers remain to realize these benefits, both in technical and social aspects of real-world adoption.

On July 27, 2021, as part of the IEEE-EMBS International Conference on Biomedical and Health Informatics jointly organized with the 17th IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks [6], a workshop was held on "Measuring Quality of Life with Multimodal Data." The workshop was divided into two sessions, the first focusing on disease detection and the second focusing on the measurement of well-being. Abstracts from the keynotes and talks are presented in Table 1; this meeting report summarizes key discussion points, relevant literature, and finally a vision for the immediate future, including how multimodal measures can provide value to stakeholders across drug development and care delivery, as well as three key areas where headway will need to be made if we are to continue to build on the encouraging progress so far: collaboration and data sharing, removal of barriers to data integration, and alignment around robust modular evaluation of new measurement capabilities.

XSL•FO
RenderX

Table 1. Talk titles and abstracts for all presented work.

Speaker	Affiliation	Title	Abstract	Citations and further reading
Session 1: disease detection				
Author LF	Evidation Health	PGHD[a]: a new ally for public health	PGHD from smartphones, wearables, and other sensors have the potential to transform the way health is measured, with broad-ranging applications from clinical research to public health and health care at large. This talk will survey examples of applications of PGHD across therapeutic areas, including post-op monitoring, screening for cognitive impairment, and a particular focus on public health applications for flu and COVID-19 detection and quantification. Finally, I will discuss lessons learned in translating PGHD research into benefits for the individual, with emphasis on the importance of evaluating analytics performance (eg, AUROC[b], sensitivity, and specificity) within a specific context of use of a real-world application.	[7,8]
Author FL	Roche Pharma Research and Early Development, F. Hoffmann-La Roche Ltd	Digital health technology tools and quality of life: examples from current studies in neurological disorders	In recent years, DHTTs[c] such as smartphones and wearables are becoming an integrated part of clinical research. Augmented by novel often AI[d]-powered signal processing, they enable continuous and precise measurements of disease symptoms. It is therefore becoming important to link these measures to the different aspects of QOL[e] of patients to make them meaningful tools for drug decision-making. In this talk, I will highlight examples from DHTTs we are developing for neurological disorders such as Parkinson disease, multiple sclerosis, and Huntington disease. Leveraging active testing and patient questionnaires accompanied by passive continuous monitoring in daily life, these tools offer rich sets of data. General signal processing and dedicated machine learning/AI solutions are used to unlock these data sets and relate them back to standard clinical scores of disease severity. I will show how resulting measures relate to patients' self-perceived health-related quality of life, how DHTTs used during COVID-19–induced lockdowns can offer new insights on QOL perception, and how we envision strengthening the link between novel sensor measurements and patient-relevant symptoms and impacts.	[9,10]
Author BV	Byteflies	Leveraging multimodal sensor data to assess complex chronic conditions at home	Byteflies's Sensor Dot platform enables continuous acquisition of physiologic and behavioral data. We leverage this multimodal data to move diagnostic tests typically performed in a specialized environment to the home of the user and to make longitudinal assessments of chronic conditions possible. In both cases, an understanding of the continuous changes in activities of daily living is crucial for safe and accurate clinical interpretation of the data. In this talk, I will discuss EpiCare@Home, a remote epileptic seizure monitoring solution built on top of the Byteflies platform.	[11,12]

XSL•FO
RenderX

Speaker	Affiliation	Title	Abstract	Citations and further reading
Author GG	Department of Neuroscience, University of Padua, Italy; SENSE-DAT Srl, Padua, Italy	Unsupervised wearable and machine learning approach to identify depression, anxiety, and stress physiological phenotypes	Background: Anxiety and depression are defined with clinical interviews in RCTs[f], possibly inflating intervention's/placebo's effects. We here introduce an algorithm to identify anxiety and depression with wearable-measured physiological biomarkers. Objectives: To validate a machine learning–based algorithm using wearable unsupervised measurements of the autonomic nervous system and physiological parameters to classify clinical anxiety and depression according to validated questionnaires. Methods: Included were physically healthy workers from the general population wearing an arm-band wearable device equipped with photoplethysmogram and electrodermal activity sensors for 24 hours. Participants answered validated self-report questionnaires for mental health, including PSS-10[g], GAD-7[h], and PHQ-9[i]. Wearable recordings were subject to artifact removal, signal preprocessing, and split in 30-second blocks for which physiological indexes and related features were extracted. A feature fusion approach was implemented together with the C5.0 machine learning algorithm, which was run on 70% randomly selected preprocessed blocks, and on the remaining 30% for external validation. Coprimary outcomes were anxiety (GAD-7≥10), and depression (PHQ-9≥10). Results: We included 95 participants (yielding 237,778 monitoring blocks), 47.7% females, mean age 37.2 (SD 15.5) years. Overall, 13.7% had anxiety, 12.6% had depression, and 7.4% had both. In the main sample, the wearable machine learning algorithm showed excellent accuracy for coprimary outcomes, namely, AUC[j]=0.928 for anxiety and AUC=0.959 for depression. Discussion: Limitations of the study include self-report questionnaires to assess primary outcomes and its cross-sectional nature. Potential implications of this work include biomarker-based inclusion criteria in RCT testing interventions for anxiety and depression, as well as screening and monitoring tools of mental health issues in the general population. Further studies should replicate the proposed algorithm against structured interview-based diagnoses with different wearable devices on clinical samples, possibly with a longitudinal design.	[13-15]
Session 2: measuring well-being				
Author AS	Rice University	Multimodal sensor data analysis and modeling for health and well-being	Digital phenotyping and machine learning technologies have shown a potential to measure objective behavioral and physiological markers, provide risk assessment for people who might have a high risk of poor health and well-being, and help make better decisions or behavioral changes to support health and well-being. I will introduce a series of studies, algorithms, and systems we have developed for measuring, predicting, and supporting personalized health and well-being. I will also discuss challenges, learned lessons, and potential future directions in health and well-being research.	[2,16]

XSL•FO
RenderX

Speaker	Affiliation	Title	Abstract	Citations and further reading
Author BS	UCSD[k] Department of Bioengineering and the Halicioglu Data Science Institute; Oura	The future of health and wellness discovery is democratic	Engineered solutions for personal data generation (wearable sensors, apps, etc) and analysis are proliferating rapidly, but health services served by these technologies continue to lag behind. Complexity in human diversity stymies algorithm generalizability and hampers successful wide adoption of any specific solution. We propose that efforts at expanding engagement in discovery will achieve two complementary goals: (1) promote mapping of biological diversity beyond demography and genetics into physiology and behavior so that algorithms can be developed on empirically determined subpopulations, and (2) fertilize natural experiments that will reveal communities sharing needs and goals, for whom solutions can then be tailored. Efforts to expand engagement may enable a virtuous cycle where iterative improvement and expansion in precision wellness technologies go from intractable to standard in personal, community, and clinical settings.	[17,18]
Author FC	Cambridge Cognition; Department of Psychiatry, University of Cambridge	Characterizing fatigue using digital technologies	Fatigue is both common and burdensome across a range of patient groups. The manifestation of fatigue is complex, comprising both subjective and objective changes to cognitive and physical performance, and is determined by a range of factors, including sleep, mood, time of day, competing demands, and environmental context, as well as disease-specific variables. These factors, and consequently the patient's experience of fatigue, vary with time, meaning that infrequent in-clinic assessments are likely to be of limited sensitivity. Given this complexity, we have been interested in exploring the potential role of digital technologies in capturing and characterizing fatigue, particularly the impact of fatigue on cognitive performance, across a range of clinical conditions. This talk will focus on methods of data collection such as brief active assessments, voice capture, and passive data from wearable technology, and describe insights these data provide us into this complex symptom.	[3,19]
Author CvH	Connected Health Solutions, imec; OnePlanet Research Center	Nanoelectronics and AI for our (and our planet's) health	We are faced with global challenges related to health, food, sustainability, and the environment. While these are formidable challenges, they also represent a substantial opportunity to improve people's lives on a global scale while at the same time creating new economic opportunities. We are convinced nanoelectronics and digital technologies are the key tools for disruptive solutions. With that purpose in mind, the OnePlanet Research Center was created as a multidisciplinary collaboration between imec, Radboud University Medical Center, and Wageningen University & Research. In OnePlanet, we apply nanoelectronics and analytics innovations to solve problems related to personalized health, personalized nutrition, mental well-being, sustainable food production, and reduced environmental impact. The sensors and data innovations are working toward the creation of digital twins for prevention, early detection, or interception of disease.	[20,21]

XSL•FO
RenderX

Speaker	Affiliation	Title	Abstract	Citations and further reading
Author SF	MIT Media Lab	Monitoring well-being using longitudinal passive data	The boundaries between the consumer and clinical device markets are becoming leaner every year. This trend is driven by a number of factors including consumer demand for ubiquitous and constantly accessible health care; increased presence of chronic conditions (eg, high blood pressure, diabetes, depression, and obesity); and a corresponding need for preventive health care, an increasingly aging global population, availability of cost-effective wearable technology, and remote access to storage and computation resources. This trend enables substantial opportunities for providing health care services to larger populations at lower cost. It will also pave the way to personalized medicine where prevention, diagnosis, and treatment of a disease can be tailored to individuals' characteristics and behavior. In this presentation, recent developments of wearable technologies at MIT Media Lab and their application to the diagnosis of mental health diseases and overall well-being are discussed.	[22,23]

[a]PGHD: person-generated health data.

[b]AUROC: area under the receiver operating characteristic curve.

[c]DHTT: digital health technology tool.

[d]AI: artificial intelligence.

[e]QOL: quality of life.

[f]RCT: randomized controlled trial.

[g]PSS-10: Perceived Stress Scale.

[h]GAD-7: Generalized Anxiety Disorder–7

[i]PHQ-9: Patient Health Questionnaire.

[j]AUC: area under the curve.

[k]UCSD: University of California, San Diego.

Background: A Shared Lexicon

To begin discussions, participants shared their perspectives on some of the terminology relevant to this emerging area of research. In Table 2, we restate some of the key points raised to orientate readers in the following report.

Table 2. Key terms relevant to the discussion. Participants shared terminology relevant to this emerging area of research.

Term	Definition	References
Multimodal measures	Referencing "Multimodal Deep Learning," multimodal measures are derived from multiple input modalities (eg, activity, sleep, heart rate, patient-reported outcomes, or contextual data)	[24]
Health-related quality of life	An individual's or a group's self-perceived physical and mental health over time	[25]
Digital measure	Sensor-derived objective measures arising from "connected digital products." Includes active tests captured via a mobile platform and continuous passive data collected from a wearable technology but excludes electronic patient-reported outcomes and other subjective measures collected from mobile platforms. An all-inclusive term, encompassing all stages of maturity, settings, and technologies.	[26,27]
Digital end point	A subset of robustly evaluated digital measures that have successfully pursued acceptance or qualification and can be used as decision-making evidence in clinical trials	[27]
Digital biomarker	Objective quantifiable physiological and behavioral data that are collected and measured by means of digital devices such as portables, wearables, implantables, or digestibles. The data collected are typically used to explain, influence, or predict health-related outcomes.	[27]
Patient-reported outcome	Assessments about how patients feel or function in their daily lives where the information is reported by the patient themselves, without interpretation or modification by someone else. Note that assessments can cover a wide range of relevant categories, some of which are more quantifiable and less subjective (including medication use or symptom presence), and some which are more subjective (including symptom severity and perception of well-being).	[26]

XSL•FO
RenderX

Session 1: Disease Detection

The first session focused on the use of multimodal data and machine learning for disease detection. Detecting deviations from normal behaviors and processes is a key step in triggering further actions, whether that be a follow-up with a health care provider or a direct digital intervention [28-30].

The session started with a keynote from author LF of Evidation Health, who spoke about how person-generated health data (PGHD; adapted from [31]) are transforming public health applications of disease detection. Author LF provided an overview of how PGHD are being used to detect and measure disease progression in a range of indications, the use of PGHD for detecting COVID-19, and the challenges of distinguishing COVID-19 from other influenza-like illnesses and infections [7]. LF underlined that machine learning model performance needs to be evaluated with a specific context of use in mind and that, without such context, model performance is ultimately of little relevance in terms of clinical utility and large-scale adoption. Finally, the keynote closed with a discussion of how what we classically think of as *evidence* can be a source of value to patients themselves, by helping them manage and understand their own health.

The following talks covered a wide range of indications, including author FL of Roche who discussed the use of smartphone-based apps to monitor neurological conditions including Parkinson disease [9] and multiple sclerosis [10,32]. Author BV shared work from the Byteflies platform showing how the system is being deployed for longitudinal monitoring of sleep disorders and cardiorespiratory and neurodegenerative conditions, and for detecting seizures in epilepsy [11,12]. Finally, author GG presented recent work examining how unsupervised measurements of autonomic nervous system signals, including photoplesmography and electrodermal activity (EDA), are showing value in the detection and staging of mental health conditions like anxiety and depression, and how these measures play a complementary role to traditional biomarkers, becoming a useful tool in enhancing clinical trials and precision psychiatry [13-15].

The session was closed with a short panel discussion featuring all the speakers that focused on questions raised by the attendees. One question addressed the pros and cons of data collection via *bring your own device* (BYOD; ie, allowing participants to connect their own devices) versus data collection via an app versus provisioned devices. The speakers agreed that there are different advantages to each approach. For example, BYOD enables comparison to a personal baseline and has advantages for device adherence, whereas provisioned devices can enable higher data uniformity and eliminate barriers to participation due to lack of access to appropriate hardware. Overall, the key is to select the right data collection approach for a given setting; where data consistency or a specific data type or density is priority, for example, in a smaller randomized controlled trial, provisioning may be preferred [33]; BYOD may in turn be preferred in settings where scale becomes limiting or where long-term "pervasive" monitoring places an emphasis on measuring ecologically valid natural behavior [34]. It was noted that while progress has been made around BYOD for patient-reported outcomes [35-37], similar progress for digital measures has not been seen and will be a key step in unlocking the value previously outlined. Another question focused on challenges to integrating objective (ie, from wearable devices) and subjective (eg, from surveys of patient-reported outcomes) inputs. The panel pointed out that many disease detection applications combine both objective and subjective inputs, for example, asking participants to confirm a signal or get a follow-up test. They also pointed out that subjective and objective inputs measure different aspects, so we should not expect them to correlate; however, this also means that they may have different relationships to a given concept of interest [38]. Thus applications that combine objective and subjective inputs can have an advantage in signal detection for disease detection. To help clarify this point, consider the following example on general well-being: a range of objective characteristics can be measured that are informative of overall well-being, including social media activity, patterns of sleep and activity, news consumption, patterns of independence, and many other objective data sources; these sources are informative of several aspects of subjective well-being (eg, perception of health), but none have a direct relationship to any specific aspect of subjective well-being, and what relationship there is differs between individuals [39].

There were also questions on the value of specific objective features (eg, EDA in stress), and GG discussed how this is a special case because this objective marker directly measures autonomic nervous system activation and thus gives a very good signal on psychological state. This was contrasted against other objective measures (eg, step counts) that have a more indirect relationship to symptoms like depression and anxiety.

The panel also discussed the impact of covariates within a cohort (eg, comorbidities) and how it influences model performance. Specifically, when trying to derive more "generalizable" models, which perform well across a broad range of unseen individuals, there is a need to incorporate a large number of covariates, and these covariates can have highly varying relevance across individuals. Progress on this topic has been made in other fields [40], but it was noted that such considerations are particularly relevant to multimodal measures.

Session 2: Measuring Well-being

The second session focused on measuring QOL and well-being. This ever-growing field has seen proof of concepts for measures across a range of health-related QOL-relevant symptoms, including fatigue [1,41], depression [3,42,43], stress [2], anxiety [44], and independence [8,45], and significant resource is being invested to understand what "wellness" means for diverse populations [46].

Author AS of Rice University started the session with a keynote on *multimodal sensor data analysis and modeling for health and well-being*, discussing her vision for how measures can underpin decision support and behavior change interventions. Her examples included schizophrenia [47], mood, and stress [16]. She also discussed challenges in in-the-wild multimodal

XSL•FO
RenderX

data modeling such as model personalization and adaptability to new users/patients and missing data.

Author BS of the University of California, San Diego then discussed current limitations in generalizability [48]. He proposed that the growth of personal sensor devices should enable us to augment classification by demographics and genetics, by including time series of physiology and behavior in our understanding of human diversity [17,18,49]. BS suggested developing algorithms that account for these dynamical differences, especially in health and wellness settings. Author FC of Cambridge Cognition then presented her work on the measurement of fatigue, which is increasingly understood to be a highly patient-relevant symptom across a large range of conditions [50-53]. She discussed the heterogeneity of the manifestations of fatigue, and their approach to combining active tests, voice biomarkers, and passive data collection to capture this complex symptom. Author CvH of imec then presented his work on digestible sensors and sensorized toilets for examining gut physiology. Finally, author SF of the Massachusetts Institute of Technology presented work on digital signals from wearables and smartphones, with application for the assessment of depression symptoms [22] and suicidal thoughts [23].

Overall, the panel discussion focused on measuring well-being as a whole versus specific aspects of QOL. While developing measures for specific aspects remains highly relevant for clinical development and treatment—for example, a measure of anxiety severity enables drug development and management of diagnosed individuals—personalized measures of general well-being have substantial application in public health and in engagement with individuals' prediagnosis. Advancements to date have focused more on the former, as the relevance to the pharmaceutical industry is higher and the validation pathway is simpler [27]. The former also limits the diversity of experience captured and so frames an opportunity for reconceptualizing wellness, health, and QOL derived from broader participation in mapping individuals' perceived needs. The panel also discussed whether it is possible and valuable to stratify mood predictions (ie, creating *semipersonalized* models where individuals with similar manifestations, personas, or journey stages are grouped together). Stratification based on objective behavioral data and digital signals can also advance our understanding of a condition by delineating commonalities across patients.

Key Discussion Points

Value of Multimodal Measures

Multimodal digital measures have expanded the number of possibilities for new ways to measure health by capturing an increasing number of proxies for multiple aspects of functions related to health. The panel emphasized that such measures are not a replacement for patient-reported outcomes but additional, complimentary tools to help understand the patients' lived experience, ideally in a low burden and unobtrusive way. The research priority should therefore focus on measures that matter when defining patients' health or general wellness. To achieve that, a 4-level sequential framework has been recently proposed by Manta et al [54] to evaluate meaningfulness of digital measures, namely, meaningful aspects of health, defining the aspect of a disease to address; specific and targeted concept of interest; outcome to be measured; and end point, including methodology and analysis plan to estimate patient improvement (eg, due to treatment).

While the majority of the research efforts are focusing on the definition and development of outcome measures, the adoption and investigation of these outcome measures in clinical trials as exploratory assessments is key to the development and validation of end points. The panel highlighted the rapidly expanding range of digital cognitive decline measures as an example and the need for the field to do more comparative studies [3] and patient-centric research [54] to focus efforts around the most meaningful and valuable candidate measures. Personalized or individual health trajectories were highlighted as potentially highly valuable, both to patients and to stakeholders outside of clinical development, for example, payer organizations exploring value-based agreements. Personalized health trajectories will require the possibility to define multiple health measures of interest, as no single measure will be equally relevant across individuals and across individual health journeys [54]. LF pointed to a key enabler being access to "healthy" data via monitoring of individuals prior to key events or diagnoses such that individualized baselines and, subsequently, individualized responses can be observed [8].

In the past decade, and accelerated by the widespread use of smartphones and other connected digital products, the use of digital products and devices in clinical trials has grown substantially, albeit primarily in observational studies and non–industry-funded clinical trials focusing on wellness [55]. The COVID-19 pandemic has by necessity further accelerated the adoption of digital health solutions for clinical research in the context of remote monitoring and telehealth [56].

Examples of the most advanced clinical applications of multimodal digital data are in Parkinson disease [57-60] and multiple sclerosis [60,61] with focus on motor function; cognitive decline in Alzheimer disease [62]; and diagnosis of depression [3], Friedreich's ataxia [63], chronic obstructive pulmonary disease [60,64], and COVID-19 [7,65,66]. Interest in multimodal digital measures is also growing among early drug discovery researchers, where personalized medicine approaches can be enabled by capturing longitudinal information on patients behaviors and in real-world settings, sometimes referred to as "digital phenotyping" [67-70]. Indeed digital measures are seen as a new component of real-world data [71]; thus to drug discovery stakeholders, multimodal measures can also serve an important role by helping to bridge the gap between evidence generation in clinical development and late-phase studies.

Challenges Remaining

As the number of technologies and sources of digital health data increases, data integration and harmonization remain open challenges. The panelists identified three key obstacles that will need to be overcome to maintain momentum in the field.

First, slow and limited collaborative efforts in prioritizing data sharing will continue to hold back at-scale development and

XSL•FO
RenderX

evaluation of novel digital measures and end points. Many companies are starting to realize the value of data sharing internally to their own walls [67], and increasingly, Findable, Accessible, Interoperable, Reusable data principles are becoming a core part of many data strategies [72]. Collaborations like the Innovative Medicines Initiative project RADAR-BASE [73-75] and the subsequent impact on a range of projects and application areas point to a possible path forward and the impact that precompetitive work in this space can have on productivity. Furthermore, multimodal sensor data is currently lacking broadly accepted and adopted common data models [76], which follow the example of other data types such as genomics and electronic health records, and have been a catalyst for progress in those fields; progress is being made [77], but more needs to be done to drive broad adoption [78]. Progress here will facilitate data integration, synchronization, and fusion that are often significant technical challenges at the individual study level when aligning and analyzing a network of connected devices [79]. A consequence of this is that substantial resources must be dedicated to technical challenges, slowing overall progress and innovation. The impact of better alignment on standards can be seen, for example, in the impact the Clinical Data Interchange Standards Consortium (CDISC) [80] has on submission data; thus it is important that collaborative efforts to make CDISC-compliant adaptations for digital health data are making progress [81]. Equally, progress on Fast Healthcare Interoperability Resources specifically on global standardization of data formats for digital health applications is encouraging [82].

Lastly, the rapid evolution of the digital health market and the short life cycle of wearables and connected devices [83,84] are challenges for data integration and reproducibility and generalization of analytical methodologies at the basis of digital measures and end points. Scaling innovation and efficient evaluation of new technologies and updated versions of hardware and software will require adherence to modular evaluation frameworks [85].

Future advances are expected from cross-industry initiatives to develop data platforms such as the Digital Medicine Society sensor integration initiative [78].

Conclusions and Path Forward

With the future of health care in mind, the panelists touched on a broad range of key takeaways. It is critical to incorporate practical, representative, and systematic approaches to involving patients in everyday health decisions [14]. Several examples discussed highlighted the importance of decision support systems or outcomes for clinical development and the value of early engagement with regulators in this space [86]. The panelists also discussed the significance in bridging the gap from measures to medicine: clinician confidence. Multimodal measures and continuous data capture are new concepts and have not been used by many practitioners, but these methods have the ability to contextualize observations and provide a direct connection to patients.

The workshop focused on sharing experiences and perspectives in the expanding use of multimodal data (multiple simultaneously collected objective data modalities, contextual information, and subjective inputs) to detect disease and capture complex outcomes. Across a wide range of examples, from infectious diseases to mental health and well-being, the speakers showcased the progress made and expressed optimism for future advancement and progression in the field.

Acknowledgments

We are grateful to the speakers who submitted their work to the workshop and took part in the panels, and to all attendees who contributed to such lively discussions. We also thank Paolo Bonato and the other conference organizers for their input and the chance to host the workshop.

Authors' Contributions

VDL and IC organized the workshop. All authors contributed to the conception and design of the article, drafted the article, and are responsible for final approval of the version to be published.

Conflicts of Interest

IC is an employee of the Digital Medicine Society. VDL is an employee of Novartis Pharma AG. BV is an employee of Byteflies. BS is a scientific advisor to and has an economic interest in OuraRing Inc. GG is an employee of the Department of Neuroscience, University of Padua Italy; he reports funding from the EU Horizon 2020–PD_Pal Grant 825785, and he owns stock in Sensedat srl. LF is cofounder of Evidation Health Inc. FL is an employee of F. Hoffmann-La Roche Ltd. AS has received travel reimbursement or honorarium payments from Gordon Research Conferences, Pola Chemical Industries, Leuven Mindgate, American Epilepsy Society, and IEEE. AS has also received research support from Microsoft, Sony Corporation, NEC Corporation, and Pola Chemicals and consulting fees from Gideon Health and Suntory Global Innovation Center. AS was paid by the European Science Foundation for a grant review.

References

1. Luo H, Lee PA, Clay I, Jaggi M, De Luca V. Assessment of fatigue using wearable sensors: a pilot study. Digit Biomark 2020;4(Suppl 1):59-72 [FREE Full text] [doi: 10.1159/000512166] [Medline: 33442581]

XSL•FO
RenderX

2. Sano A, Taylor S, McHill AW, Phillips AJ, Barger LK, Klerman E, et al. Identifying objective physiological markers and modifiable behaviors for self-reported stress and mental health status using wearable sensors and mobile phones: observational study. J Med Internet Res 2018 Jun 08;20(6):e210 [FREE Full text] [doi: 10.2196/jmir.9410] [Medline: 29884610]
3. Sverdlov O, Curcic J, Hannesdottir K, Gou L, De Luca V, Ambrosetti F, et al. A study of novel exploratory tools, digital technologies, and central nervous system biomarkers to characterize unipolar depression. Front Psychiatry 2021;12:640741. [doi: 10.3389/fpsyt.2021.640741] [Medline: 34025472]
4. Abbas A, Sauder C, Yadav V, Koesmahargyo V, Aghjayan A, Marecki S, et al. Remote digital measurement of facial and vocal markers of major depressive disorder severity and treatment response: a pilot study. Front Digit Health 2021;3:610006 [FREE Full text] [doi: 10.3389/fdgth.2021.610006] [Medline: 34713091]
5. Vandendriessche B, Godfrey A, Izmailova ES. Multimodal biometric monitoring technologies drive the development of clinical assessments in the home environment. Maturitas 2021 Sep;151:41-47. [doi: 10.1016/j.maturitas.2021.06.009] [Medline: 34446278]
6. IEEE BHI-BSN-2021. URL: https://www.bhi-bsn-2021.org/ [accessed 2021-08-03]
7. Shapiro A, Marinsek N, Clay I, Bradshaw B, Ramirez E, Min J, et al. Characterizing COVID-19 and influenza illnesses in the real world via person-generated health data. Patterns (N Y) 2021 Jan 08;2(1):100188 [FREE Full text] [doi: 10.1016/j.patter.2020.100188] [Medline: 33506230]
8. Karas M, Marinsek N, Goldhahn J, Foschini L, Ramirez E, Clay I. Predicting subjective recovery from lower limb surgery using consumer wearables. Digit Biomark 2020;4(Suppl 1):73-86 [FREE Full text] [doi: 10.1159/000511531] [Medline: 33442582]
9. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
10. Midaglia L, Mulero P, Montalban X, Graves J, Hauser SL, Julian L, et al. Adherence and satisfaction of smartphone- and smartwatch-based remote active testing and passive monitoring in people with multiple sclerosis: nonrandomized interventional feasibility study. J Med Internet Res 2019 Aug 30;21(8):e14863 [FREE Full text] [doi: 10.2196/14863] [Medline: 31471961]
11. Dan J, Vandendriessche B, Paesschen WV, Weckhuysen D, Bertrand A. Computationally-efficient algorithm for real-time absence seizure detection in wearable electroencephalography. Int J Neural Syst 2020 Nov;30(11):2050035. [doi: 10.1142/S0129065720500355] [Medline: 32808854]
12. Swinnen L, Chatzichristos C, Jansen K, Lagae L, Depondt C, Seynaeve L, et al. Accurate detection of typical absence seizures in adults and children using a two-channel electroencephalographic wearable behind the ears. Epilepsia 2021 Nov;62(11):2741-2752. [doi: 10.1111/epi.17061] [Medline: 34490891]
13. Fava M, Evins AE, Dorer D, Schoenfeld D. The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. Psychother Psychosom 2003;72(3):115-127. [doi: 10.1159/000069738] [Medline: 12707478]
14. Fernandes BS, Williams LM, Steiner J, Leboyer M, Carvalho AF, Berk M. The new field of 'precision psychiatry'. BMC Med 2017 Apr 13;15(1):80 [FREE Full text] [doi: 10.1186/s12916-017-0849-x] [Medline: 28403846]
15. Carvalho AF, Solmi M, Sanches M, Machado MO, Stubbs B, Ajnakina O, et al. Evidence-based umbrella review of 162 peripheral biomarkers for major mental disorders. Transl Psychiatry 2020 May 18;10(1):152. [doi: 10.1038/s41398-020-0835-5] [Medline: 32424116]
16. Li B, Sano A. Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. Proc ACM Interactive Mobile Wearable Ubiquitous Technologies 2020 Jun 15;4(2):1-26. [doi: 10.1145/3397318]
17. Smarr BL, Ishami AL, Schirmer AE. Lower variability in female students than male students at multiple timescales supports the use of sex as a biological variable in human studies. Biol Sex Differ 2021 Apr 22;12(1):32 [FREE Full text] [doi: 10.1186/s13293-021-00375-2] [Medline: 33888158]
18. Smarr BL, Aschbacher K, Fisher SM, Chowdhary A, Dilchert S, Puldon K, et al. Feasibility of continuous fever monitoring using wearable devices. Sci Rep 2020 Dec 14;10(1):21640. [doi: 10.1038/s41598-020-78355-6] [Medline: 33318528]
19. Cormack FK, Barnett JH, Taptiklis N. Acoustic features of voice as a measure of cognitive load during performance of serial subtraction in a remote data collection context. Alzheimer Dementia 2021 Dec 31;17(S11):e056271. [doi: 10.1002/alz.056271]
20. van Kraaij AWJ, Schiavone G, Lutin E, Claes S, Van Hoof C. Relationship between chronic stress and heart rate over time modulated by gender in a cohort of office workers: cross-sectional study using wearable technologies. J Med Internet Res 2020 Sep 09;22(9):e18253 [FREE Full text] [doi: 10.2196/18253] [Medline: 32902392]
21. Smets E, Rios Velazquez E, Schiavone G, Chakroun I, D'Hondt E, De Raedt W, et al. Large-scale wearable data reveal digital phenotypes for daily-life stress detection. NPJ Digit Med 2018;1:67. [doi: 10.1038/s41746-018-0074-9] [Medline: 31304344]

XSL•FO
RenderX

22. Ghandeharioun A, Fedor S, Sangermano L, Ionescu D, Alpert J, Dale C, et al. Objective assessment of depressive symptoms with machine learning and wearable sensors data. 2017 Presented at: Seventh International Conference on Affective Computing and Intelligent Interaction; October 23-26, 2017; San Antonio, TX. [doi: 10.1109/acii.2017.8273620]
23. Kleiman EM, Turner BJ, Fedor S, Beale EE, Picard RW, Huffman JC, et al. Digital phenotyping of suicidal thoughts. Depress Anxiety 2018 Jul;35(7):601-608. [doi: 10.1002/da.22730] [Medline: 29637663]
24. Ngiam J, Khosla A, Kim M, Nam J, Lee H, Ng AY. Multimodal deep learning. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. 2011 Presented at: ICML'11; June 28-July 2, 2011; Bellevue, WA p. 689-696 URL: https://dl.acm.org/doi/10.5555/3104482.3104569
25. Health-related quality of life (HRQOL). Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/hrqol/index.htm [accessed 2021-12-17]
26. Coravos A, Goldsack JC, Karlin D, Nebeker C, Perakslis E, Zimmerman N, et al. Digital medicine: a primer on measurement. Digit Biomark 2019;3(2):31-71 [FREE Full text] [doi: 10.1159/000500413] [Medline: 32095767]
27. Goldsack JC, Dowling AV, Samuelson D, Patrick-Lake B, Clay I. Evaluation, acceptance, and qualification of digital measures: from proof of concept to endpoint. Digit Biomark 2021;5(1):53-64 [FREE Full text] [doi: 10.1159/000514730] [Medline: 33977218]
28. Brönneke JB, Debatin JF, Hagen J, Kircher P, Matthies H, Stern AD. DiGA VADEMECUM: How to Launch Digital Health Apps in the German Healthcare System. Berlin: MWV; 2021.
29. Maricich YA, Bickel WK, Marsch LA, Gatchalian K, Botbyl J, Luderer HF. Safety and efficacy of a prescription digital therapeutic as an adjunct to buprenorphine for treatment of opioid use disorder. Curr Med Res Opin 2021 Feb;37(2):167-173 [FREE Full text] [doi: 10.1080/03007995.2020.1846022] [Medline: 33140994]
30. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
31. Determining real-world data's fitness for use and the role of reliability. Margolis Center for Health Policy. 2019. URL: https://healthpolicy.duke.edu/sites/default/files/2019-11/rwd_reliability.pdf [accessed 2021-08-23]
32. Montalban X, Graves J, Midaglia L, Mulero P, Julian L, Baker M, et al. A smartphone sensor-based digital outcome assessment of multiple sclerosis. Mult Scler 2022 Apr;28(4):654-664 [FREE Full text] [doi: 10.1177/13524585211028561] [Medline: 34259588]
33. Mueller A, Hoefling HA, Muaremi A, Praestgaard J, Walsh LC, Bunte O, et al. Continuous digital monitoring of walking speed in frail elderly patients: noninterventional validation study and longitudinal clinical trial. JMIR Mhealth Uhealth 2019 Nov 27;7(11):e15191 [FREE Full text] [doi: 10.2196/15191] [Medline: 31774406]
34. Kaye J. Public Policy Aging Rep 2017;27(2):53-61 [FREE Full text] [doi: 10.1093/ppar/prx005] [Medline: 31148911]
35. Gwaltney C, Coons SJ, O'Donohoe P, O'Gorman H, Denomey M, Howry C, et al. "Bring Your Own Device" (BYOD): the future of field-based patient-reported outcome data collection in clinical trials? Ther Innov Regul Sci 2015 Nov;49(6):783-791. [doi: 10.1177/2168479015609104] [Medline: 30222388]
36. Byrom B, Gwaltney C, Slagle A, Gnanasakthy A, Muehlhausen W. Measurement equivalence of patient-reported outcome measures migrated to electronic formats: a review of evidence and recommendations for clinical trials and bring your own device. Ther Innov Regul Sci 2019 Jul;53(4):426-430. [doi: 10.1177/2168479018793369] [Medline: 30157687]
37. Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value Health 2018 May;21(5):581-589 [FREE Full text] [doi: 10.1016/j.jval.2017.10.008] [Medline: 29753356]
38. Houts CR, Patrick-Lake B, Clay I, Wirth RJ. The path forward for digital measures: suppressing the desire to compare apples and pineapples. Digit Biomark 2020;4(Suppl 1):3-12 [FREE Full text] [doi: 10.1159/000511586] [Medline: 33442577]
39. Voukelatou V, Gabrielli L, Miliou I, Cresci S, Sharma R, Tesconi M, et al. Measuring objective and subjective well-being: dimensions and data sources. Int J Data Sci Analytics 2020 Jun 29;11(4):279-309. [doi: 10.1007/s41060-020-00224-2]
40. de Jong VMT, Moons KGM, Eijkemans MJC, Riley RD, Debray TPA. Developing more generalizable prediction models from pooled studies and large clustered data sets. Stat Med 2021 Jul 10;40(15):3533-3559 [FREE Full text] [doi: 10.1002/sim.8981] [Medline: 33948970]
41. IDEA-FAST. URL: https://idea-fast.eu/ [accessed 2021-08-03]
42. Makhmutova M, Kainkaryam R, Ferreira M, Min J, Jaggi M, Clay I. Prediction of self-reported depression scores using person-generated health data from a virtual 1-year mental health observational study. In: Proceedings of the 2021 Workshop on Future of Digital Biomarkers. 2021 Presented at: DigiBiom '21; June 25, 2001; Virtual p. 4-11. [doi: 10.1145/3469266.3469878]
43. Jacobson NC, Weingarden H, Wilhelm S. Digital biomarkers of mood disorders and symptom change. NPJ Digit Med 2019;2:3. [doi: 10.1038/s41746-019-0078-0] [Medline: 31304353]
44. Jacobson NC, Summers B, Wilhelm S. Digital biomarkers of social anxiety severity: digital phenotyping using passive smartphone sensors. J Med Internet Res 2020 May 29;22(5):e16875 https://www.jmir.org/2020/5/e16875/. [doi: 10.2196/16875] [Medline: 32348284]

XSL•FO
RenderX

45. Bahej I, Clay I, Jaggi M, De Luca V. Prediction of patient-reported physical activity scores from wearable accelerometer data: a feasibility study. In: Masia L, Micera S, Akay M, Pons JL, editors. Converging Clinical and Engineering Research on Neurorehabilitation III: Proceedings of the 4th International Conference on NeuroRehabilitation (ICNR2018), October 16-20, 2018, Pisa, Italy. Cham: Springer; 2019:668-672.
46. All of Us Research Program. 2020. URL: https://allofus.nih.gov/future-health-begins-all-us [accessed 2021-12-17]
47. Tseng VW, Sano A, Ben-Zeev D, Brian R, Campbell AT, Hauser M, et al. Using behavioral rhythms and multi-task learning to predict fine-grained symptoms of schizophrenia. Sci Rep 2020 Sep 15;10(1):15100. [doi: 10.1038/s41598-020-71689-1] [Medline: 32934246]
48. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digit Health 2020 Sep;2(9):e489-e492 [FREE Full text] [doi: 10.1016/S2589-7500(20)30186-2] [Medline: 32864600]
49. Grant AD, Newman M, Kriegsfeld LJ. Ultradian rhythms in heart rate variability and distal body temperature anticipate onset of the luteinizing hormone surge. Sci Rep 2020 Nov 23;10(1):20378. [doi: 10.1038/s41598-020-76236-6] [Medline: 33230235]
50. Elefante E, Tani C, Stagnaro C, Ferro F, Parma A, Carli L, et al. Impact of fatigue on health-related quality of life and illness perception in a monocentric cohort of patients with systemic lupus erythematosus. RMD Open 2020 Feb;6(1):e001133 [FREE Full text] [doi: 10.1136/rmdopen-2019-001133] [Medline: 32075876]
51. Rodríguez Antolín A, Martínez-Piñeiro L, Jiménez Romero ME, García Ramos JB, López Bellido D, Muñoz Del Toro J, et al. Prevalence of fatigue and impact on quality of life in castration-resistant prostate cancer patients: the VITAL study. BMC Urol 2019 Oct 16;19(1):92 [FREE Full text] [doi: 10.1186/s12894-019-0527-8] [Medline: 31619215]
52. McCabe RM, Grutsch JF, Braun DP, Nutakki SB. Fatigue as a driver of overall quality of life in cancer patients. PLoS One 2015;10(6):e0130023 [FREE Full text] [doi: 10.1371/journal.pone.0130023] [Medline: 26070133]
53. Eaton-Fitch N, Johnston SC, Zalewski P, Staines D, Marshall-Gradisnik S. Health-related quality of life in patients with myalgic encephalomyelitis/chronic fatigue syndrome: an Australian cross-sectional study. Qual Life Res 2020 Jun;29(6):1521-1531 [FREE Full text] [doi: 10.1007/s11136-019-02411-6] [Medline: 31970624]
54. Manta C, Patrick-Lake B, Goldsack JC. Digital measures that matter to patients: a framework to guide the selection and development of digital measures of health. Digit Biomark 2020;4(3):69-77 [FREE Full text] [doi: 10.1159/000509725] [Medline: 33083687]
55. Marra C, Chen JL, Coravos A, Stern AD. Quantifying the use of connected digital products in clinical research. NPJ Digit Med 2020;3:50. [doi: 10.1038/s41746-020-0259-x] [Medline: 32285011]
56. Marra C, Gordon WJ, Stern AD. Use of connected digital products in clinical research following the COVID-19 pandemic: a comprehensive analysis of clinical trials. BMJ Open 2021 Jun 22;11(6):e047341 [FREE Full text] [doi: 10.1136/bmjopen-2020-047341] [Medline: 34158302]
57. Viceconti M, Hernandez Penna S, Dartee W, Mazzà C, Caulfield B, Becker C, et al. Toward a regulatory qualification of real-world mobility performance biomarkers in Parkinson's patients using digital mobility outcomes. Sensors (Basel) 2020 Oct 20;20(20):5920 [FREE Full text] [doi: 10.3390/s20205920] [Medline: 33092143]
58. Morgan C, Craddock I, Tonkin E, Kinnunen K, McNaney R, Whitehouse S, et al. Protocol for PD SENSORS: Parkinson's disease symptom evaluation in a naturalistic setting producing outcome measures using SPHERE technology. An observational feasibility study of multi-modal multi-sensor technology to measure symptoms and activities of daily living in Parkinson's disease. BMJ Open 2020 Nov 30;10(11):e041303 [FREE Full text] [doi: 10.1136/bmjopen-2020-041303] [Medline: 33257491]
59. Merola A, Sturchio A, Hacker S, Serna S, Vizcarra JA, Marsili L, et al. Technology-based assessment of motor and nonmotor phenomena in Parkinson disease. Expert Rev Neurother 2018 Nov;18(11):825-845. [doi: 10.1080/14737175.2018.1530593] [Medline: 30269610]
60. Rochester L, Mazzà C, Mueller A, Caulfield B, McCarthy M, Becker C, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the Mobilise-D approach. Digit Biomark 2020;4(Suppl 1):13-27 [FREE Full text] [doi: 10.1159/000512513] [Medline: 33442578]
61. Woelfle T, Pless S, Wiencierz A, Kappos L, Naegelin Y, Lorscheider J. Practice effects of mobile tests of cognition, dexterity, and mobility on patients with multiple sclerosis: data analysis of a smartphone-based observational study. J Med Internet Res 2021 Nov 18;23(11):e30394 [FREE Full text] [doi: 10.2196/30394] [Medline: 34792480]
62. Muurling M, de Boer C, Kozak R, Religa D, Koychev I, Verheij H, RADAR-AD Consortium. Remote monitoring technologies in Alzheimer's disease: design of the RADAR-AD study. Alzheimers Res Ther 2021 Apr 23;13(1):89 [FREE Full text] [doi: 10.1186/s13195-021-00825-4] [Medline: 33892789]
63. Mueller A, Paterson E, McIntosh A, Praestgaard J, Bylo M, Hoefling H, et al. Digital endpoints for self-administered home-based functional assessment in pediatric Friedreich's ataxia. Ann Clin Transl Neurol 2021 Sep;8(9):1845-1856. [doi: 10.1002/acn3.51438] [Medline: 34355532]
64. Rahman MJ, Nemati E, Rahman M, Vatanparvar K, Nathan V, Kuang J. Toward early severity assessment of obstructive lung disease using multi-modal wearable sensor data fusion during walking. Annu Int Conf IEEE Eng Med Biol Soc 2020 Jul;2020:5935-5938. [doi: 10.1109/EMBC44109.2020.9176559] [Medline: 33019325]

XSL•FO
RenderX

65. Ni X, Ouyang W, Jeong H, Kim J, Tzaveils A, Mirzazadeh A, et al. Automated, multiparametric monitoring of respiratory biomarkers and vital signs in clinical and home settings for COVID-19 patients. Proc Natl Acad Sci U S A 2021 May 11;118(19):e2026610118 [FREE Full text] [doi: 10.1073/pnas.2026610118] [Medline: 33893178]
66. Quer G, Radin JM, Gadaleta M, Baca-Motes K, Ariniello L, Ramos E, et al. Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 2021 Jan;27(1):73-77. [doi: 10.1038/s41591-020-1123-x] [Medline: 33122860]
67. Hartl D, De Luca V, Kostikova A, Laramie J, Kennedy S, Ferrero E, et al. Translational precision medicine: an industry perspective. J Transl Med 2021 Jun 05;19(1):245 [FREE Full text] [doi: 10.1186/s12967-021-02910-6] [Medline: 34090480]
68. Insel T. Digital phenotyping: technology for a new science of behavior. JAMA 2017 Oct 03;318(13):1215-1216. [doi: 10.1001/jama.2017.11295] [Medline: 28973224]
69. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019;2:88. [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
70. Fagherazzi G. Deep digital phenotyping and digital twins for precision health: time to dig deeper. J Med Internet Res 2020 Mar 03;22(3):e16770 [FREE Full text] [doi: 10.2196/16770] [Medline: 32130138]
71. Dagenais S, Russo L, Madsen A, Webster J, Becnel L. Use of real-world evidence to drive drug development strategy and inform clinical trial design. Clin Pharmacol Ther 2022 Jan;111(1):77-89. [doi: 10.1002/cpt.2480] [Medline: 34839524]
72. Wise J, de Barron AG, Splendiani A, Balali-Mood B, Vasant D, Little E, et al. Drug Discov Today 2019 Apr;24(4):933-938 [FREE Full text] [doi: 10.1016/j.drudis.2019.01.008] [Medline: 30690198]
73. RADAR-base. GitHub. URL: https://github.com/RADAR-base [accessed 2022-02-07]
74. RADAR-base. 2018. URL: https://radar-base.org/ [accessed 2022-02-07]
75. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Hyve, RADAR-CNS Consortium. RADAR-Base: open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
76. OMOP Common Data Model. Observational Health Data Sciences and Informatics. URL: https://www.ohdsi.org/data-standardization/the-common-data-model/ [accessed 2021-12-17]
77. IEEE Standard for Open Mobile Health Data--representation of metadata, sleep, and physical activity measures. IEEE Standards Association. URL: https://standards.ieee.org/standard/1752_1-2021.html [accessed 2021-12-21]
78. Clay I, Angelopoulos C, Bailey AL, Blocker A, Carini S, Carvajal R, et al. Sensor data integration: a new cross-industry collaboration to articulate value, define needs, and advance a framework for best practices. J Med Internet Res 2021 Nov 09;23(11):e34493 [FREE Full text] [doi: 10.2196/34493] [Medline: 34751656]
79. Mazzà C, Alcock L, Aminian K, Becker C, Bertuletti S, Bonci T, et al. Technical validation of real-world monitoring of gait: a multicentric observational study. BMJ Open 2021 Dec 02;11(12):e050785 [FREE Full text] [doi: 10.1136/bmjopen-2021-050785] [Medline: 34857567]
80. Facile R, Muhlbradt EE, Gong M, Li Q, Popat V, Pétavy F, et al. Use of Clinical Data Interchange Standards Consortium (CDISC) standards for real-world data: expert perspectives from a qualitative Delphi survey. JMIR Med Inform 2022 Jan 27;10(1):e30363 [FREE Full text] [doi: 10.2196/30363] [Medline: 35084343]
81. Hume S, Aerts J, Sarnikar S, Huser V. Current applications and future directions for the CDISC Operational Data Model standard: a methodological review. J Biomed Inform 2016 Apr;60:352-362 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.016] [Medline: 26944737]
82. Lehne M, Luijten S, Vom Felde Genannt Imbusch P, Thun S. The use of FHIR in digital health - a review of the scientific literature. Stud Health Technol Inform 2019 Sep 03;267:52-58. [doi: 10.3233/SHTI190805] [Medline: 31483254]
83. Gurova O, Merritt T, Papachristos E, Vaajakari J. Sustainable solutions for wearable technologies: mapping the product development life cycle. Sustainability 2020 Oct 14;12(20):8444. [doi: 10.3390/su12208444]
84. Brönneke JB, Müller J, Mouratis K, Hagen J, Stern AD. Regulatory, legal, and market aspects of smart wearables for cardiac monitoring. Sensors (Basel) 2021 Jul 20;21(14):4937 [FREE Full text] [doi: 10.3390/s21144937] [Medline: 34300680]
85. Goldsack JC, Coravos A, Bakker J, Bent B, Dowling A, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55. [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
86. Cerreta F, Ritzhaupt A, Metcalfe T, Askin S, Duarte J, Berntgen M, et al. Digital technologies for medicines: shaping a framework for success. Nat Rev Drug Discov 2020 Sep;19(9):573-574. [doi: 10.1038/d41573-020-00080-6] [Medline: 32398879]

Abbreviations

BYOD: bring your own device
CDISC: Clinical Data Interchange Standards Consortium
EDA: electrodermal activity
PGHD: person-generated health data
QOL: quality of life

XSL•FO
RenderX

Edited by T Leung; submitted 23.12.21; peer-reviewed by J Curcic, H Luderer, D Sato, B Booth; comments to author 04.02.22; revised version received 14.02.22; accepted 25.04.22; published 26.05.22.

Please cite as:
Clay I, Cormack F, Fedor S, Foschini L, Gentile G, van Hoof C, Kumar P, Lipsmeier F, Sano A, Smarr B, Vandendriessche B, De Luca V
Measuring Health-Related Quality of Life With Multimodal Data: Viewpoint
J Med Internet Res 2022;24(5):e35951
URL: https://www.jmir.org/2022/5/e35951
doi:10.2196/35951
PMID:35617003

©Ieuan Clay, Francesca Cormack, Szymon Fedor, Luca Foschini, Giovanni Gentile, Chris van Hoof, Priya Kumar, Florian Lipsmeier, Akane Sano, Benjamin Smarr, Benjamin Vandendriessche, Valeria De Luca. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2022. 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.

XSL•FO
RenderX

Viewpoint

Digital Transformation of Face-To-Face Focus Group Methodology: Engaging a Globally Dispersed Audience to Manage Institutional Change at the World Health Organization

Gaya Gamhewage[1], MBBS, MPH, MSc; Mohamed Essam Mahmoud[1], MBBS, MSc; Anna Tokar[1], MSc, PhD; Melissa Attias[1], MSc; Christos Mylonas[1], BA, MA, PhD; Sara Canna[1], MA; Heini Utunen[1], B Soc Sc, MA

World Health Organization, Genève, Switzerland

Corresponding Author:
Gaya Gamhewage, MBBS, MPH, MSc
World Health Organization
Av Appia 20
Genève, 1211
Switzerland
Phone: 41 794755563
Email: gamhewageg@who.int

Abstract

Focus group discussions (FGDs) are widely used to obtain qualitative data from purposely selected groups of people. This paper describes how the Learning and Capacity Development (LCD) unit of the World Health Organization (WHO) Health Emergencies Programme (WHE) digitalized FGDs to engage with WHO staff from around the world, to listen, share, and collect their feedback in the development of a WHO learning framework. The impact of the COVID-19 pandemic and the introduction of local lockdowns and travel restrictions resulted in the wide use of digital platforms, such as Zoom, for employee communications and collaboration capable of reaching employees wherever they are working. The LCD/WHE team drew upon the experience of WHO colleagues from human resources, country, and regional offices to set up and hold FGDs in 6 languages with participants from all WHO regions. Building on the findings of a 2019 WHO staff survey, which was part of a comprehensive, organization-wide career development initiative, the digitalized FGDs allowed for the exchange of substantive feedback, novel ideas, and alignment, connecting across different geographies, disciplines, and levels of seniority. As a result, FGDs can be successfully conducted online, but it is essential to remove barriers to participation by adopting a multilingual and flexible approach in multinational and international organizations such as the WHO.

(J Med Internet Res 2022;24(5):e28911) doi:10.2196/28911

KEYWORDS

qualitative research; digitalization; WHO; World Health Organization; FGDs; focus group discussions

Introduction

As part of its transformation process to meet the health challenges of the 21st century, the World Health Organization (WHO) is developing the first-ever global Learning Strategy for health personnel around the world. This Learning Strategy was initiated by the WHO Academy project to provide strategic direction for the operations of the WHO Academy, as well as to frame the broader strategic vision in the domain of learning to achieve health goals that the WHO's Member States and stakeholders could use as a framework for the future. The WHO Academy is a transformative project to revolutionize lifelong learning in health and is currently being established. Based on this global Learning Strategy, a learning framework is being developed to ensure the upskilling and reskilling of all WHO staff and contractors, establish the norm of lifelong learning, and help transform the WHO into a learning institution. In combination with literature reviews, this process necessitates desk reviews, surveys, and consultations, as well as more in-depth qualitative research on staff views, visions, and suggestions on the "what" and "how" of the learning framework's role in accompanying them along new career pathways. Learning is in fact one of the key enabling factors that facilitates staff in their career and professional development and constitutes a key requirement for staff to remain up-to-date, relevant, and skilled to perform certain technical roles at their best. The focus group discussions (FGDs) aimed at engaging staff in their career development by identifying the learning needs that they have in relation to their career goals and the

XSL•FO
RenderX

challenges they face in acquiring or enhancing certain skills that are critical to advance along a chosen career pathway.

Furthermore, due to the pandemic, it was essential that FGDs be digitized in light of lockdowns and travel restrictions. Therefore, the Learning and Capacity Development unit of the WHO Health Emergencies Programme (WHE) designed a streamlined methodology that ensured staff from all 6 WHO regions and headquarters (HQ) provide their perspectives and perceptions. This paper describes the methodology and the lessons learned from the digitalization of the FGDs to reach WHO staff in all 6 regions and HQ to be used in other contexts where qualitative research is carried out digitally for geographically dispersed populations.

Methodology

Designing the FGDs

Based on qualitative research approaches in the health care guidebook [1], a core group of WHE personnel with qualitative research expertise was established to lead the process. The first step was to develop the key principles that would yield the broadest possible participation from a wide range of WHO staff by overcoming already identified barriers. These barriers included geographical dispersion and disengagement from corporate transformation projects; poor internet access; language; and social barriers, such as perceptions of power differentials that may hinder those from the WHO's country and regional offices from participating; as well as concerns related to the confidentiality and legitimacy of the process itself and its impact toward meaningful change at the WHO. The second step involved scaling up the human resources required to run a large number of FGDs in multiple languages over a limited time frame of 7 days. The third step involved identifying and training a larger research team to ensure that sufficiently robust, high-quality FGDs could be conducted. The final step in the design involved the development of the tools and process for the FGDs, a quality-control mechanism, and a support system for the research team.

Expanding and Training the Research Team

A total of 27 personnel from the WHE, human resources (HR), and regional offices volunteered to participate as facilitators, notetakers, and hosts for the FGDs. All volunteers were required to attend some 1-hour training sessions to prepare them for their roles. The training aimed to improve facilitation skills focusing on working online and equip them with the necessary digital tools. The facilitators were native or advanced-level Arabic, English, French, Portuguese, Russian, and Spanish speakers.

In order to standardize all sessions, the facilitators were trained to use the FGD script (Multimedia Appendix 1). The FGD script was based on the WHO FGD guidance [2], with questions formulated to reflect the insights gained from a WHO survey conducted in mid-2019 on staff perceptions of career development and learning. The script was tested and then further adapted after holding 2 pilot FGDs in English and French with members of the organizing team. This allowed for collecting feedback from participants and identifying possible bottlenecks including those related to technology failures. In addition, in each FGD session, a notetaker was assigned as "a silent observer" supporting the facilitator by providing notes on various practical aspects that could hinder the smooth running of the FGDs (eg, internet connectivity challenges, which could impact engagement, body language, key messages; Multimedia Appendix 2).

Inviting and Enrolling the Participants

The research team invited WHO staff to participate in the FGDs using a convenience sampling method [1]; invitations with a complete description of the project were sent via corporate emails with a link to a sign-up form. In addition, as it was essential to create legitimacy, confidence, and trust in the FGDs, the WHO HQ research team collaborated with the corporate HR team to craft the appropriate communication messages to invite staff to participate.

The invitation included a personalized video message by the research team lead outlining the purpose of the survey, highlighting the importance of staff participation, and making a firm commitment by the team to preserving confidentiality and using data appropriately. The combination of formal and personalized invitations by a senior staff member offered increased motivation to participants to enroll in the exercise.

Conducting the FGD Sessions

All FGDs were conducted through the online videoconferencing platform Zoom. Gender and geographical balance were ensured whenever possible. The sessions started with a plenary meeting where all FGD participants received the same introductions and were then sent into breakout rooms according to their language or group preferences. Standard scripts were used for the plenary and breakout rooms to ensure consistency. The duration of each FGD was approximately 60 minutes. Verbal informed consent was obtained from every participant at the start of the FGDs.

Processing of FGD Data

The research team video recorded, translated into English, and transcribed all FGDs. All data were uploaded into NVivo (version 12; QSR International). The video recordings were secured for transcription and then deleted for confidentiality reasons, and the identity of the participants was kept anonymous. A full narrative report of the findings was produced. A member with the role of checking, as envisaged by the validation technique used in qualitative research [3], was introduced. The findings were presented at an all-staff seminar to check for accuracy and resonance with staff experiences.

Ethics Considerations

An ethics review was not applicable for this study because this paper is based on an internal consultation process in the WHO. The consent of all the participants was requested and obtained at the start of each online FGD, and their consent was recorded.

Results

The participatory approach yielded positive results, with 401 staff enrolling in the study within 5 days, of a total of 8000 WHO staff. Those who signed up to participate were split into

XSL•FO
RenderX

groups based on self-identification according to the following criteria:

1. Priority groups (National Professional Officers; women in/seeking a leadership role; young professionals; and general service staff in secretarial, administrative, and logistic functions)
2. Language preference (Arabic, Chinese, English, French, Portuguese, Russian, or Spanish)
3. WHO region (African, Americas, Eastern Mediterranean, European, South-East Asian, Western Pacific, or HQ)
4. Time preference (morning or afternoon Central European Time).

The priority groups were identified from the findings of the 2019 mixed methods survey on WHO staff career development and learning. An individual's presence in 1 priority group did not preclude them from belonging to others, and indeed, many participants identified themselves as belonging to more than 1 category.

A total of 180 participants were available to participate during the 7 days set for the FGDs. In total, 38 FGD sessions in 7 languages were conducted, with 5 participants on average in each group. Although 45% (180/401) of those who enrolled actually participated, staff from all WHO regions and HQ were represented. In most groups, some participants could not use the video function due to low bandwidth, so observation of body language was limited. However, all were able to use audio.

The findings were arranged as follows:

- General findings across all groups, with the following categories related to learning for staff: expectations, priority transversal skills for all staff, technical or job-specific skills, perceived enablers, perceived disablers
- Specific findings related to learning priorities (general service staff, National Professional Officers, women in/seeking a leadership role, young professionals under 40 years of age)
- Ideas related to the WHO Academy (expectations, priority learning activities, fears)
- Links to the new WHO career pathways initiative (expectations, fears)

To validate the findings, nearly 500 staff from all WHO regions participated in the all-staff seminar where the summary findings were presented. No comments were received that challenged the summary findings. The use of anonymized quotes was described by staff as being powerful. Staff commented that the process of participation in the FGDs was motivating in itself, and for many, this was the first time they felt "heard and seen by colleagues in other parts of the organization," especially at the global level; it was a learning exercise to hear others' views and perspectives; it yielded socially positive results ("felt great" or "connected" or "as part of one family") and they would be happy to participate in future FGDs; it led to an increased willingness to use the methodology online for other purposes; and it resulted in requests to create a forum for the participants to stay connected going forward beyond participating in the FGDs.

The findings were used to revise the first-ever WHO global Learning Strategy, make recommendations for the WHO Academy and the elaboration of learning and career pathways, and develop a learning framework to support staff progress. The methodology will now be used as a standard methodology in the WHE and in other WHO staff engagement initiatives to gain the perspectives of key stakeholders for strategy, program, and policy development in the future. By-products such as an online forum initiated by the participants of the first digital FGDs for staff to stay engaged are also underway.

Discussion

The WHE designed a qualitative study using online FGDs that ensured staff from all 6 WHO regions and HQ participated to provide their perspectives and perceptions to support establishing the global Learning Strategy for the WHO Academy and to support the elaboration of the learning pathways as a key component of career development. In this paper, we argue that despite the many stated challenges of conducting online qualitative research, FGDs can be successfully conducted online. Many researchers, especially social researchers, faced multiple challenges to continue their face-to-face interactions and fieldwork due to public health security measures imposed by governments worldwide since the start of the COVID-19 pandemic [4]. As a result, a digital and nondigital range of ideas and methods were trialed to continue fieldwork in pandemic times [5]. However, much research was conducted online in previous years, and many examples of online surveys, interviews, and digital ethnographies are available in the literature [6-8].

To successfully digitalize FGDs, it is essential to remove barriers to participation by adopting a multilingual and flexible approach in multinational and international organizations such as the WHO, where staff have busy schedules and are separated geographically and hierarchically. These results are consistent with those of other studies and support the digitalization of interviews and FGDs as the most used qualitative methods [9-11].

Additionally, online FGDs have the potential, when designed with consideration of the organizational and participatory contexts, to yield rich results in the form of eliciting not only knowledge but also sentiment. They have collateral advantages of helping personnel in a dispersed organization to feel more connected with each other and be more seen and heard by the power centers of an organization as well as by peers in other locations, with the positive consequence of generating staff engagement. They offer new means of influencing significant change and strategies of a global organization. These social benefits align with the concept that we are currently experiencing a social age characterized by a less hierarchical structure, participation in problem identification, and cocreation of creative and contextualized solutions, rather than command and control of the power centers of an organization [12].

The research team's decision to engage all staff allowed us to capture rich and varied ideas, thoughts, opinions, and lived experiences that gave voice to employees' needs and aspirations across different regions and positions. Moreover, such an

XSL•FO
RenderX

approach could contribute to creating trusting relationships and building rapport, and thus could decrease possible information bias. This methodological approach allowed meaningful conversations to take place, recognizing each participant's active role in the process of knowledge cocreation and, by so doing, increasing equity. We also believe that triangulation could diminish researchers' bias, which was achieved by collecting data from different facilitators and notetakers. Additionally, we were guided by an emergent research design, which consists in considering the whole process as an iterative cycle in which the preliminary findings of the first FGDs informed the subsequent ones. The key lessons learned can be summarized as follows:

1. Designing the FGDs
   - Set the most relevant values and principles as foundations for the design of the online FGDs and explicitly link them to the larger processes of clarifying meaning and significance
   - Proactively overcome barriers including physical, social, institutional, and psychological barriers (time zones, language, geographical distance, equity of access in participation, trust, credibility, meaning)
2. Preparing the team
   - Train facilitators on competencies for running FGDs and on using the technology
   - Do a test run—test methodology and technical tools
   - Standardize the tools—formal training and unified scripts
3. Running the FGDs
   - Run daily debriefing sessions for the facilitation team and offer facilitator support to answer questions and provide coaching
   - Ensure equity by inviting all who signed up—even if there were hundreds, and even if it means innovating and expanding the FGD rollout plan
   - Deal with low bandwidth—cameras off when necessary
   - Keep to time—do not inconvenience participants
   - Plan for the worst-case scenario—have alternate staff available for facilitator, notetaker, and host roles
   - Be proactive—send reminders to attendees and staff before the start of the event
   - Be flexible—participants and staff may be late, and a group may need to be rescheduled
   - Anticipate reductions in turnout, even among confirmed participants, providing an opportunity for them to join another FGD
   - Do not assume digital literacy or familiarity with selected tech platforms or tools—the more explanation, the better
   - Consider how social cues are different online—the awkwardness of knowing when to speak and difficulty observing body language
   - Report back to participants and all other stakeholders while maintaining confidentiality
   - Integrate findings concretely into ongoing processes
   - Provide results in multiple formats that are targeted to different audiences
4. Expanding the benefits
   - Appreciate the facilitation team—provide coaching and certificates
   - Use the process to keep personnel engaged in major change initiatives
   - Use the methodology beyond research to engage stakeholders, to gain feedback on programs, and in planning
   - Capitalize and empower the use of other tools and digital fora to maximize social benefits—a strong sense of community, a sense of contributing to something meaningful, and having a voice

Conclusions

FGDs can be successfully conducted online. To effectively digitalize FGDs, it is essential to remove barriers to participation by adopting a multilingual and flexible approach. Online FGDs have the potential to yield rich results in the form of eliciting not only knowledge but also sentiment and capturing rich and varied ideas, thoughts, opinions, and lived experiences. This methodological approach allowed for meaningful conversations, recognizing the role of each participant in the process of knowledge cocreation and promotion of equity.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group discussion script for WHO staff learning.
[DOCX File , 19 KB - jmir_v24i5e28911_app1.docx]

Multimedia Appendix 2
Focus group discussion notes and registration form for WHO staff learning.
[DOCX File , 15 KB - jmir_v24i5e28911_app2.docx]

References

1. Pope C, Mays N. Qualitative Research in Health Care 3rd Edition. Oxford: Blackwell/BMJ; Jan 2006.

XSL•FO
RenderX

2. WHO safe childbirth checklist collaboration: evaluation report. World Health Organization. URL: https://apps.who.int/iris/rest/bitstreams/1093997/retrieve [accessed 2021-01-19]
3. Birt L, Scott S, Cavers D, Campbell C, Walter F. Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 2016 Jun 22;26(13):1802-1811. [doi: 10.1177/1049732316654870] [Medline: 27340178]
4. Howlett M. Looking at the 'field' through a Zoom lens: methodological reflections on conducting online research during a global pandemic. Qual Res 2021 Jan 16. [doi: 10.1177/1468794120985691]
5. Lupton D. Doing fieldwork in a pandemic (crowd-sourced document). 2021 Jul 05. URL: https://docs.google.com/document/d/1clGjGABB2h2qbduTgfqribHmog9B6P0NvMgVuiHZCl8/edit# [accessed 2021-04-01]
6. Beaulieu A. Mediating ethnography: objectivity and the making of ethnographies of the internet. Soc Epistemol 2004 Apr;18(2-3):139-163. [doi: 10.1080/0269172042000249264]
7. Jenner BM, Myers KC. Intimacy, rapport, and exceptional disclosure: a comparison of in-person and mediated interview contexts. Int J Soc Res Methodol 2018 Aug 22;22(2):165-177. [doi: 10.1080/13645579.2018.1512694]
8. Lo Iacono V, Symonds P, Brown DH. Skype as a tool for qualitative research interviews. Sociol Res Online 2016 May 31;21(2):103-117. [doi: 10.5153/sro.3952]
9. Hooley T, Marriott J, Wellens J. What is Online Research? Using the Internet for Social Science Research. London: Bloomsbury Publishing; Jun 21, 2012:1-118.
10. James N, Busher H. Online Interviewing. London: SAGE Publications Ltd; 2009.
11. Cowan J. Br J Educ Technol 2014 Oct 20;45(6):E30-E30 [FREE Full text] [doi: 10.1111/bjet.12225_6]
12. Stodd J. The Social Leadership Handbook Second Edition. Bournemouth: Sea Salt Learning; Jun 25, 2016.

Abbreviations

FGD: focus group discussion
HQ: headquarters
HR: human resources
LCD: Learning and Capacity Development unit
WHE: WHO Health Emergencies Programme
WHO: World Health Organization

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by A Mahnke, B Ploderer, MDG Pimentel; comments to author 08.06.21; revised version received 28.02.22; accepted 20.04.22; published 26.05.22.

Please cite as:
Gamhewage G, Mahmoud ME, Tokar A, Attias M, Mylonas C, Canna S, Utunen H
Digital Transformation of Face-To-Face Focus Group Methodology: Engaging a Globally Dispersed Audience to Manage Institutional Change at the World Health Organization
J Med Internet Res 2022;24(5):e28911
URL: https://www.jmir.org/2022/5/e28911
doi:10.2196/28911
PMID:35617007

©Gaya Gamhewage, Mohamed Essam Mahmoud, Anna Tokar, Melissa Attias, Christos Mylonas, Sara Canna, Heini Utunen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2022. 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.

XSL•FO
RenderX

Viewpoint

Types of Errors Hiding in Google Scholar Data

Romy Sauvayre[1,2], PhD, HDR

[1]Laboratoire de Psychologie Sociale et Cognitive, Centre national de la recherche scientifique, Université Clermont Auvergne, Clermont-Ferrand, France

[2]Polytech Clermont, Clermont Auvergne INP, Aubière, France

Corresponding Author:
Romy Sauvayre, PhD, HDR
Polytech Clermont
Clermont Auvergne INP
2 avenue Blaise Pascal
TSA 60 026
Aubière, 63178
France
Phone: 33 473405545
Fax: 33 956577681
Email: romy.sauvayre@uca.fr

Abstract

Google Scholar (GS) is a free tool that may be used by researchers to analyze citations; find appropriate literature; or evaluate the quality of an author or a contender for tenure, promotion, a faculty position, funding, or research grants. GS has become a major bibliographic and citation database. For assessing the literature, databases, such as PubMed, PsycINFO, Scopus, and Web of Science, can be used in place of GS because they are more reliable. The aim of this study was to examine the accuracy of citation data collected from GS and provide a comprehensive description of the errors and miscounts identified. For this purpose, 281 documents that cited 2 specific works were retrieved via Publish or Perish software (PoP) and were examined. This work studied the false-positive issue inherent in the analysis of neuroimaging data. The results revealed an unprecedented error rate, with 279 of 281 (99.3%) examined references containing at least one error. Nonacademic documents tended to contain more errors than academic publications (U=5117.0; P<.001). This viewpoint article, based on a case study examining GS data accuracy, shows that GS data not only fail to be accurate but also potentially expose researchers, who would use these data without verification, to substantial biases in their analyses and results. Further work must be conducted to assess the consequences of using GS data extracted by PoP.

(J Med Internet Res 2022;24(5):e28354) doi:10.2196/28354

KEYWORDS

reference accuracy; database reliability; false positives; academic publication; research evaluation; scientometrics; citation analysis

Introduction

Google Scholar (GS) has become a major bibliographic and citation database. Soon after its creation in 2004, GS received major criticism [1], but subsequently, further studies described it more positively [2,3]. Indeed, the literature acknowledges the free access offered by GS [3-5] and the quality of its coverage [6-12]. The coverage of GS is considered better than that of both Web of Science (WoS) [12-15] and Scopus [9,10], which are GS's fee-based competitors. This is particularly true regarding its coverage of social sciences and humanities research [10,16,17], conference proceedings [10,14], and books [17]. The GS database has been substantially qualitatively [18] and quantitatively [10,19] improved in all scientific areas such that, according to de Winter et al [18], it could supplant WoS.

However, "the automatic indexing of GS inevitably causes many errors" [20], such as duplicates [21] and false-positive citations [18]. Most researchers generally claim that these errors are negligible [9,10,12,20,22-24], whereas others consider that data cleaning is necessary [16,19,25] but laborious [4,21]. Thus, some scholars have used GS without data cleaning [2,6,11,14,26,27], while others have identified and removed duplicates [4,9,12,17-19,24,28,29]. This removal was performed in 23 of 36 studies (41.8%) using GS data. Furthermore, compared to the authors of related studies, these researchers less frequently identified false positives [17,18,30,31], missing values or omission errors [20,23], document type errors [18,32],

XSL•FO
RenderX

author name errors [18,33], publication year errors [16,18,33], title errors [18,33], URL errors [16,32], citation miscounts [32], and inaccessible document errors [30]. None of these 36 studies mentioned any verification of journal names in their data cleaning process. Nevertheless, Haddaway et al [4] attempted to explain the causes of duplicates, showing that they arise from typographical and capitalization errors occurring in journal names. Their findings were confirmed by a study conducted by Valderrama-Zurián et al [34] based on Scopus data.

However, an analysis of 36 articles published between 2008 and 2018 in journals with an impact factor from Journal Citation Reports (JCR) collected from WoS, GS, and relevant studies cited in the most cited research in this field showed that the data verification was not systematically followed by the calculation and reporting of an error rate. Indeed, 14 of the 36 (38.9%) studies explicitly indicated the number or rate of errors. A median error rate of 14.6% with a range from 0.04% to 53.5%, among corpora of citations ranging in size between 127 and 183,596, was calculated. Note that for those studies that were missing error rates but nevertheless had reported adequate results, the error rates were calculated and included. In addition, these studies reported error data of a median of only 1 type of error (range 0-6), and duplicates represented the error type most frequently searched for in this sample of literature (23 of 36).

This median error rate therefore demonstrates that errors are recurrent in GS data. However, GS is a free tool that may be used by researchers to analyze citations; find appropriate literature [35,36]; or evaluate the quality or influence [37] of an author or a contender for tenure, promotion, a faculty position, funding, or research grants [1,21]. Thus, the more an author is cited in a field, the more likely that person is to be considered a highly qualified researcher [38,39]. GS may also be used in research evaluations [23]. Thus, a comprehensive study of this failure of GS may be useful to the scientific community and researchers who want to use this database, whatever their field of study. However, as far as can be seen, no study reports and meticulously quantifies the different types of errors encountered in the GS data extracted by Publish or Perish software (PoP), even though such a study would allow (1) better identification of the limitations of studies based on these data, as described by Hicks et al [40] in the context of research evaluation; (2) enrichment of the thoughtful methodological reflection on potential exposure to GS errors; and (3) development of appropriate methods to limit the negative effects of GS errors on the results produced.

This case study aimed to examine the GS data extracted by PoP, provide a full count of the errors contained in the collected data, and present an epistemological reflection. By doing so, this study offers detailed categorizations of GS data that have not been provided by previous studies. The purpose is especially to address the following questions: (1) What types of GS errors could affect the data and results of researchers' studies? (2) What methodological problems may result from these errors? (3) How reliable can the citations of GS be without data cleaning?

Methods

Context

This GS study is part of broader research that aims to explore the diffusion process of neuroimaging work that sought to alert the scientific community to the issue of false positives. Two references were examined. The first reference is a poster presented at the 15th Annual Meeting for the Organization for Human Brain Mapping (OHBM) [41], and the second is an article published in the short-lived *Journal of Serendipitous and Unexpected Results* (JSUR) [42]. The question was which researchers contributed to this diffusion or, in other words, who cited the OHBM poster or the JSUR article. The collection of citation data became necessary. Nevertheless, some full texts of the citing documents collected by GS did not cite either the OHBM poster or the JSUR article. Thus, this case study was conceived. The reliability of GS data needed to be quantified to identify the limitations of the results produced with GS data before using these data in the diffusion study. This categorization of errors using these 2 references enables one to identify how GS works with literature not referenced by journal editors' websites. GS uses "automated software, known as *parsers*, to identify bibliographic data" [43] of documents available on the internet. Then, the parser software "typically" collects the same data from full documents without metadata, as the 2 references used in this case study.

Data Collection

To examine the reliability and accuracy of GS, the citations of both the OHBM poster and JSUR article were analyzed. Note that GS was the only citation database available to collect the citation data for these 2 works because neither WoS nor Scopus indexed them.

PoP version 5 was used to extract references that cited the poster. According to Harzing [44], this software provides a perfect collection of GS data ("Publish or Perish is as accurate or as inaccurate as Google Scholar itself"). In addition, PoP is a common tool in scientometric studies using GS data [14,19,29,45].

The citation data were then collected from GS via PoP. The first author's name ("Bennett, Craig M") was entered without quotation marks, and the first part of the OHBM poster and JSUR article title ("Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon") was entered with quotation marks into the "All of the words" software query box. As PoP's manual explains, the "All of the words" query "matches the search terms anywhere in the searched documents (author, title, source, abstract, references, etc)" [46], as GS does. Thus, this query was used to reproduce the same request with PoP and GS.

This title is so specific that only the following 2 results appeared: (1) 127 references cited the JSUR article [42], and (2) 154 references cited the OHBM poster [41]. In contrast, the reference that appeared in PoP and on GS was a paper supposedly published in a supplement of the famous *NeuroImage* journal and indexed by ScienceDirect. In reality, *NeuroImage* did not publish a journal article written by Bennett et al in 2009 about

XSL•FO
RenderX

the neuroimaging work. This *NeuroImage* paper does not exist. What this supplemental issue of *NeuroImage* does contain is the program of the OHBM conference. Therefore, when the citing documents cite the "*NeuroImage* ghost paper," they actually cite the OHBM poster.

Note that the JSUR article title is almost identical to the OHBM poster title—only the term "proper" in the second part of the title differs. The advantage of this strong similarity is the ability to evaluate the capacity of GS to manage citations of similar references.

A total of 281 references were extracted via PoP on October 6, 2017. Two CSV files were obtained (Multimedia Appendix 1), one for each neuroimaging reference. In this study, several columns that contained the following information were examined: authors, title of the citing document, publication year, publication or source, publisher, and web address of the citing document ("Article URL" as provided by GS). Each column was manually verified, and inaccuracies were counted and categorized in the following 6 steps:

1. The full text of accessible citing documents was downloaded and recorded.
2. The reference list of each citing document was consulted to verify and record the presence of the neuroimaging reference (OHBM poster, JSUR article, or both).
3. The document type was determined and recorded by reading it and searching for additional information on its source.
4. For each citing document, an accurate reference was elaborated for use as a standard and to determine whether GS data contain errors. An inductive and descriptive methodological approach was used to list and identify all the error types that occurred in the GS data. The reference accuracy literature served as a guide to avoid omitting the important errors in this field. A typology was elaborated and presented in the results section as follows: (1) Data collection errors (duplicates, reprints, translations, missing URLs, and inaccessible documents); (2) Academic publication collection errors (retrieval of types of documents other than journal articles, books, book chapters, and conference proceedings); (3) Citation errors (false positives or citation counted by GS when the document does not cite the reference counted); (4) Author errors (missing authors, added authors, missing part of the author's name, and errors in initials); (5) Title errors (incorrect or incomplete title, and spelling or typographical errors); (6) Publication year errors (erroneous or missing date of publication); (7) Publication of source errors (journal name errors identified in the "Publication" column of GS); and (8) Publisher errors (book editor name errors identified in the "Publisher" column of GS).
5. The GS errors found in each extracted column were listed.
6. The identified errors were aggregated by reference.

This collection, verification, and aggregation process required approximately 170 hours of work.

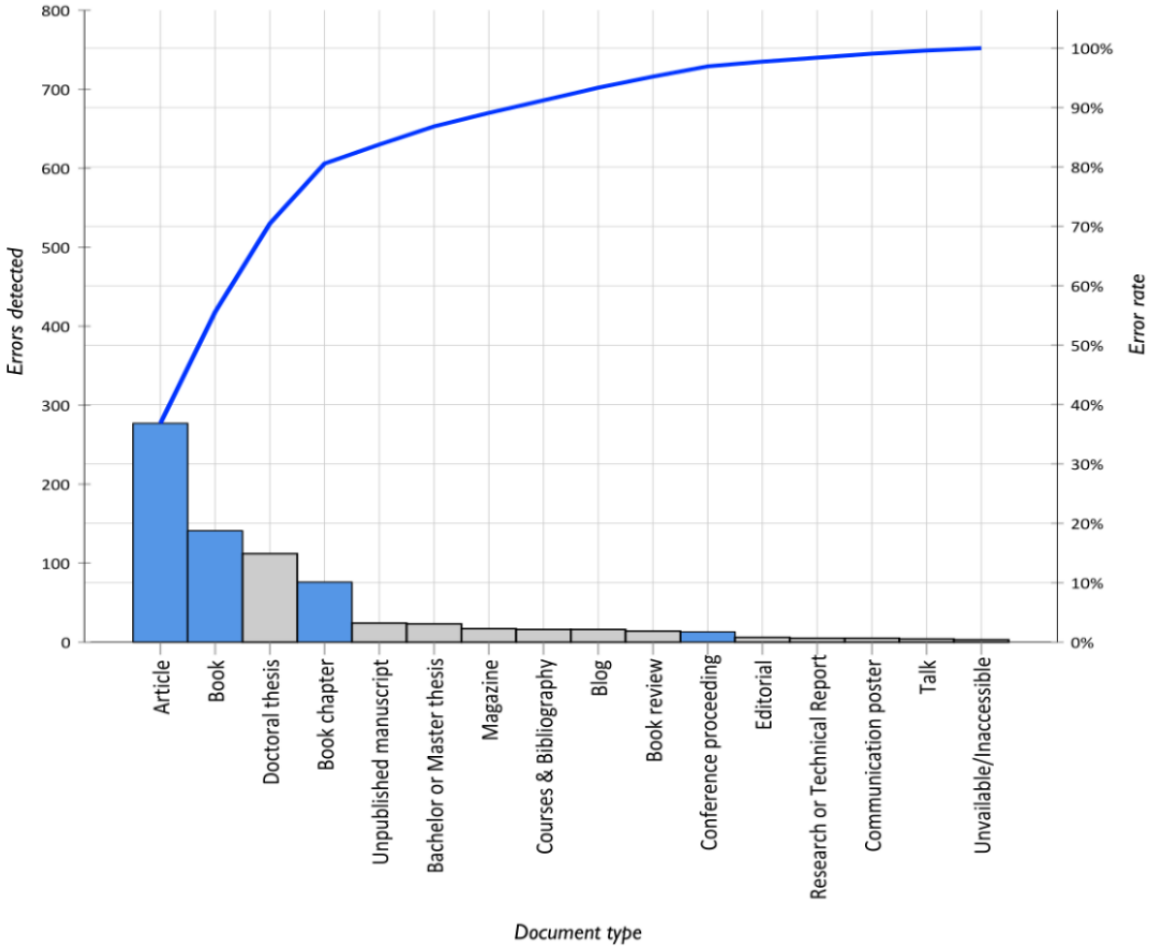
Results

Number of Errors

A total of 755 errors were detected in 281 references retrieved from GS (Figure 1), for an average of 2.7 errors (range 0-7) (Multimedia Appendix 2). Furthermore, 279 of 281 (99.3%) references contained at least one error.

XSL•FO
RenderX

Figure 1. Pareto diagram. Sum of errors detected (N=755) as a function of document type. Light blue indicates academic publications, gray indicates nonacademic documents, and dark blue indicates cumulative sum curve of errors detected.



Typology of GS Errors

After a manual examination of the references extracted from GS, the following 8 types of errors were identified (Table 1): (1) Data collection errors; (2) Academic publication collection errors; (3) Citation errors (false positives); (4) Author errors; (5) Title errors; (6) Publication year errors; (7) Publication errors; and (8) Publisher errors.

Table 1. Typology of Google Scholar errors. Typology and proportion of errors identified as a function of the number of valid references examined and as a function of the total number of errors detected.

Error type	Errors identified, n (%)	
	As a function of the number of valid references examined (n=271-281)	As a function of the total number of errors detected (N=755)
Data collection	42 (5.6)	33 (11.7)
Academic publication collection	77 (10.2)	77 (27.5)
Citation	81 (10.7)	81 (29.9)
Author	61 (8.1)	53 (19.4)
Title	60 (7.9)	57 (20.8)
Publication year	31 (4.1)	31 (11.3)
Publication	155 (20.5)	133 (47.5)
Publisher	248 (32.8)	244 (86.8)

Data Collection Errors

Data collection errors included duplicates, reprints, translations, missing URLs, and inaccessible documents (Multimedia Appendix 3). This type of error was identified in 33 of 281 (11.7%) references, and among these errors, duplicates were detected in 16 of 281 (5.7%) references. In addition, URL analysis indicated that none of the GS data in any of the PoP extractions contained duplicate URLs. However, because 18

XSL•FO
RenderX

URLs were missing, a manual search for these references was conducted to obtain and verify them. Among these missing URL references, only 2 of the 18 citing documents were inaccessible, and 9 references were duplicates, translations, or reprints.

Academic Publication Collection Errors

Some scientometric studies have used document type as a variable. Consequently, some researchers have focused exclusively on journal articles [3,6,29,30,47,48], whereas others have presented their collected citations per document type, including journal articles, books, book chapters, and conference proceedings [25]. Furthermore, "grey literature" [4], such as theses and research reports [18], can also be included. Considering the diversity of this research method, it will be interesting to further explore the document types that GS is likely to retrieve and count.

GS describes itself as a database that "provides a simple way to broadly search for scholarly literature" [49]. However, what does "scholarly literature" mean for GS? The definition provided by GS, and used in the document inclusion process, encompasses "journal papers, conference papers, technical reports or their drafts, dissertations, preprints, postprints, or abstracts" [43]. On another webpage, GS mentions that users "can search across many disciplines and sources: articles, theses, books, abstracts, and court opinions" [49]. By contrast, GS excludes "news or magazine articles, book reviews, and editorials" [43] because they are "not appropriate" [43]. Nevertheless, there is no statement about the rejection of these undesirable documents from the GS index.

In this study, the document type of each reference collected from GS was examined to determine whether the document in question was an "academic publication." In this way, a document was considered an "academic publication" only if it was (1) an article that was published in a journal with an International Standard Serial Number (ISSN) or (2) a book, a book chapter, or conference proceedings published with an International Standard Book Number (ISBN). All other document types (thesis, magazine, communication poster, bibliography, course, report, and unpublished document), so-called "nonacademic documents," were classified as GS collection errors. Note that, according to this definition, a doctoral thesis is an academic work but not an academic publication.

As Multimedia Appendix 4 shows, GS retrieved 203 of 281 (72.5%) academic publications, but included 77 nonacademic documents in the corpus. The error rate reached 27.5% according to the definition given in the literature, whereas the GS definition led to a lower error rate (6.8%). In addition, because GS data are asymmetrically distributed, a nonparametric (Mann-Whitney) test was conducted with SPSS 25 (IBM Corp), and it revealed that the nonacademic documents tended to contain more errors than the academic publications (U=5117.0; P<.001) (Multimedia Appendix 5).

Citation Errors (False Positives)

The reference list of each citing document was examined to determine which of the 2 references (OHBM poster or JSUR article) had been cited. A total of 271 full documents were available and were read. This assessment revealed that 81 of the documents (29.9%) did not cite the reference retrieved from GS. In other words, 29.9% of the citations counted by GS were false positives. In 8 of the 271 (3.0%) cases, neither of the 2 references were found. In 12 of the 271 (4.4%) cases, the JSUR article reference was found instead of the OHBM poster reference, that is, in the extraction of citations attributed by GS to the OHBM poster. Conversely, in 61 of the 271 (22.5%) cases, the OHBM poster reference was found instead of the JSUR article reference.

Additionally, these citation errors (false positives) affected 8 times more OHBM poster references than JSUR article references (odds ratio 7.77, 4.4 < CI < 13.71). Note that the OHBM poster reference was misreferenced in the citing documents more often than the JSUR article reference.

Author Errors

As Multimedia Appendix 6 shows, 53 of 273 (19.4%) references contained at least one author error. For example, initials were removed or added. Authors were missing in 41 of the 273 (15.0%) references. Surprisingly, they were replaced by a journal name or by the title of either their own book or their own book chapter. Finally, 104 authors were missing, while 20 authors were improperly added. In summary, 124 of 565 authors (22.0%) were inaccurate.

Title Errors

A thorough examination of the "Title" column extracted from GS showed that 57 of 274 (20.8%) references contained at least one error (Multimedia Appendix 7). The incompleteness of the title was the most common error identified. As a result of this error, some incomplete titles were similar to other publication titles. Furthermore, several errors were more questionable, such as replacement of a book title with a chapter title from the aforementioned book or with the title of a different chapter from another book by an author who contributed a chapter to this book. Other questionable title errors were the assemblage of 2-chapter parts published in the same book and the replacement of the publication title by its editor's name or the domain name of the website that hosts it. Surprisingly, irrelevant parts were added to the publication title, such as an ISBN number, the price of the book, the name of the book collection, and an excerpt from the front page of a thesis ("a dissertation submitted for the degree of Doctor of Philosophy"). Lastly, the reference titles also contained typographical or spelling errors.

Publication Year Errors

Publication year errors were detected in 31 of 274 (11.3%) references. In most cases, the years were missing (they were replaced by "zero" in 22 references). In other cases, the actual publication year of the JSUR article or the OHBM poster was increased by 1 year or decreased by 1, 3, 7, or 100 years (Multimedia Appendix 8).

Publication of Source Errors

The "Publication" or "Source" column retrieved from GS via PoP showed inconsistencies that depended on the document type of references (Multimedia Appendix 9). Indeed, it contained journal names, books, edited book titles, conference proceeding titles, magazine names, publisher names, domain names of

XSL•FO
RenderX

websites that host the citing documents, irrelevant parts of references, and even an author's address. Furthermore, a large number of missing values (ie, "not provided" in Multimedia Appendix 9) were found in these publication data, affecting 1 in 3 (32.0%) references. These missing publication data were observed most often for theses (bachelor's, master's, and doctoral theses) and book references.

In total, 133 of 280 (47.5%) references contained errors Table 2). These errors were mostly identified in conference proceedings, edited books, and journal articles. In addition, only half of the citations counted by GS were usable as academic publication material (Table 2, "Utility" column) because, for example, GS provides a domain name instead of the academic journal name. Among these usable data, 133 were inaccurate. Finally, in this corpus, only 60 of 280 (21.4%) references had proper usable data.

Table 2. Accurate and inaccurate content identified in the "Publication" column retrieved from Google Scholar via Publish or Perish (N=280).

Type of error	Inaccurate publication, n (%)	Accurate publication, n (%)	Utility[a]
Journal name (n=108)	56 (51.9)	52 (48.1)	(+)
Magazine name (n=2)	1 (50.0)	1 (50.0)	(−)
Book title (n=13)	13 (100.0)	0 (0.0)	(−)
Edited book title (n=29)	21 (72.4)	8 (27.6)	(+)
Conference proceeding title (n=5)	5 (100.0)	0 (0.0)	(+)
Thesis title (n=2)	2 (100.0)	0 (0.0)	(−)
Publisher name (n=2)	2 (100.0)	0 (0.0)	(−)
Domain name (n=20)	18 (90.0)	2 (10.0)	(−)
Preprint database name (n=4)	1 (25.0)	3 (75.0)	(−)
Other (n=5)	5 (100.0)	0 (0.0)	(−)
Missing value (not provided) (n=90)	9 (100.0)	81 (90.0)	(−)
Total (n=280)	133 (47.5)	147 (52.5)	N/A[b]

[a]The usable publication content for studies using academic publications is denoted by "+." The errors were not easy to categorize because of nonacademic documents. For instance, when the document type is a blog post or an unpublished draft, a journal name is not expected in the "Publication" column and thus is counted as an inaccuracy. Nevertheless, this type of document had already been counted as a data collection error. Therefore, each document type was specifically analyzed to avoid falsely increasing the error count. However, the categorization was easier for other references, such as when the journal editor name was provided instead of the journal name. In addition, an examination of spelling and typographical errors, including capitalization errors, was conducted.

[b]N/A: not applicable.

These source inconsistencies mainly occurred in journal names as typographical errors, particularly capitalization errors (Multimedia Appendix 10). The second most frequent error was title and journal name incompleteness. Journal names were heavily truncated, as shown in the following examples: "Journal of …" instead of "Journal of Advertising Research" and "Rev …" instead of "Revista de neurologia." The same type of inaccuracy was identified in the edited book titles as follows: "… Routledge Handbook of …" instead of "The Routledge Handbook of Neuroethics" and "… Imaging of the …" instead of "Imaging of the Pelvis, Musculoskeletal System, and Special Applications to CAD." Furthermore, as several journal names begin with "Journal of" and several edited books begin with "Routledge Handbook of," the incompleteness of GS data may cause difficulties.

Publisher Errors

The "Publisher" column retrieved from GS provided a variety of content (Multimedia Appendix 11) as follows: editor name (including journal editor), journal name, domain name of the website that hosts the citing document (eg, 42 of the domain names were "books.google.com"), digital library (ie, JSTOR), and missing values. The "Publisher" column contained the highest error rate found in the GS data, which was 244 of 281 (86.8%) references (Table 3). Indeed, the 248 inaccuracies detected in this column constituted a third (32.9%) of the total errors identified. Journal editors and domain names were frequently inaccurate. The utility of this publisher data was then limited to studies using academic publication data. Only the editor names of books, book chapters, and conference proceedings were usable, but they actually represented 35 of the 281 (12.5%) references. Furthermore, an error rate of 37.1% was found in these usable data. For example, an editor's name was replaced by an irrelevant name (The Penguin Press by Australia Books and Palgrave Macmillan by Springer).

XSL•FO
RenderX

Table 3. Accurate and inaccurate content in the "Publisher" column retrieved from Google Scholar via Publish or Perish (N=281).

Type of error	Inaccurate publication, n (%)	Accurate publication, n (%)	Utility[a]
Book and conference proceeding editor (n=35)	13 (37.1)	22 (62.9)	(+)
Journal editor (n=51)	51 (100.0)	0 (0.0)	(−)
Journal name (n=1)	1 (100.0)	0 (0.0)	(−)
Digital library name (n=2)	2 (100.0)	0 (0.0)	(−)
Domain name (n=167)	167 (100.0)	0 (0.0)	(−)
Not provided (n=25)	10 (40.0)	15 (60.0)	(−)
Total (n=281)	244 (86.8)	37 (13.2)	N/A[b]

[a]The usable publication content for studies using academic publication data is denoted by "+."

[b]N/A: not applicable.

Discussion

Principal Findings

The aim of this study was to examine the accuracy of citation data collected from GS via PoP and to provide a comprehensive description of the errors and miscounts identified. In fact, the extraction of raw data with inaccuracies from GS may generate incorrect results in several research areas, such as bibliometrics, scientometrics, and research evaluation. Despite the data cleaning performed by researchers (mainly duplicate removal), citation counts retrieved from GS were generally used without substantial caution. Furthermore, few comprehensive studies listed the different types of GS errors, and no previous research seemed to quantify the inherent problems of GS citations collected by PoP. This study was therefore conducted to provide a meticulous analysis of GS data to anticipate the risk of errors that may affect the data and the results of studies using them.

Ranking of GS Errors

The GS errors were analyzed using 281 documents that cited a neuroimaging work performed to raise awareness of false-positive results in the scientific community. This study revealed an unprecedented error rate, with 279 of 281 (99.3%) examined references containing at least one error. Academic publications were not free from errors. They accounted for 503 of the 755 (67.0%) detected errors. However, nonacademic documents tended to contain more errors than academic publications (U=5117.0; P<.001).

The cumulative error rate detected in this study (99.3% of references containing at least one error) differs from the median rate (14.6%) reported in the literature over the past 10 years. This difference may be explained by several aspects of previous research. First, an automatic approach was generally used to clean the data, while a manual examination was conducted in this study. Second, a varied but low number of variables were examined in these studies. A median of 1 type of error was examined in previous studies, while 8 types of errors were examined in this study. Third, the usual purpose of these studies was to compare the coverage of GS, WoS, and Scopus; thus, the researchers mainly verified duplicates in an aggregated corpus drawn from these 3 databases. Fourth, these studies did not cumulate the number of errors identified per reference.

These discrepancies make comparison difficult, but data provided by de Winter et al [18] ("Online Supplementary Material 5 Excel File") make it possible. Through these data, an error rate cumulated by reference was calculated to compare what is comparable. However, as these researchers used 4 error types, the comparison was performed for academic publication collection errors, author errors, title errors, and duplicates. All other things being equal, this study reports an error rate 3 times higher than that reported by de Winter et al [18] (64.8% and 20.5%, respectively). These findings suggest that citation counts and references extracted from GS are not fully reliable and may expose the researchers who use them to numerous errors. Note that the content of GS is the result of automatic indexing of websites by robots. The coverage depends on the indexed websites. Moreover, according to GS, "robots generally try to index every paper from every website they visit, including most major sources and also many lesser known ones" [50]. Thus, the reliability of GS is a type of "photography" of the reliability of authors' and editors' websites. Since errors can happen, it is important to identify the possible impact of GS's lack of reliability with respect to research data.

The Impact of GS Errors in Research Data

What is the probable impact of GS errors in the citation analysis or research evaluation area when citation counts and references are used without data cleaning?

Publisher Errors

The useful content that a researcher needs to find in the "Publisher" column extracted from GS via PoP is the editor name for books, book chapters, and conference proceedings. However, this column mainly contains the domain name of the website hosting the citing document. Thus, only 7.8% of these collected data are free from errors and are usable in an academic publication study. The "Publication" column therefore requires meticulous examination before use. The first step is to determine the document type of each collected reference because GS still does not provide it.

Publication Errors

The PoP manual indicates that the "Publication" column contains "journal name or similar," and "similar" is not explicitly defined, which is "not always available" and "sometimes wrong" [51]. However, the "publication" content

XSL•FO
RenderX

is more disparate and incorrect than this. Indeed, it contains journal names, book titles, thesis titles, publisher names, and domain names, and one-third of this column involves missing values. Only 21.4% of "publication" data are free from errors and are usable in an academic publication study. In addition, GS errors can impact studies in the following ways. First, the GS error rate can negatively affect the evaluation of journal impact factors and the journal ranking. Second, missing values (32.4% of references) can alter relational database management [52]. Third, typographical errors (including capitalization errors) can lead to duplicates [4]. Lastly, note that there is a risk in using the "publication" data because of the large number of errors detected in the journal names, edited book titles, and conference proceeding titles.

Citation Errors (False Positives)

The GS citation count is distorted by documents that do not cite the reference retrieved. This point is often reported in the literature, for instance, as a "phantom citation" or "false citation" [20], but no reported error rate [18,30,32,53,54] is as high as the rate found in this study (29.9%). This difference can be explained by the highly similar titles of the 2 references examined (OHBM poster and JSUR article). This finding also demonstrates the difficulties of addressing this type of similarity in GS data. Consequently, researchers may use data samples that contain false-positive citations and then may obtain biased findings.

Academic Publication Collection Errors

GS failed to retrieve only academic publications. Indeed, 27.5% of the citing documents were nonacademic publications, including doctoral theses, magazine articles, preprints, reports, courses, bibliographies, and blog posts. This error rate confirms previous findings [30]. However, if the GS definition of "scholarly literature" is applied, this error rate falls to 6.8%. This GS definition differs widely from the definition of "academic publication" used in this study. Thus, GS seems to inaccurately report citation counts and references of academic publications, and consequently, it does not accurately reflect the dissemination of published work. Therefore, the results of many scientometric studies using GS data to examine the publication activity of scientists, particularly in research evaluation, may be questionable when these data are not verified and cleaned (document types and false positives). The citation counts and h-index scores calculated by GS are also questionable. This raises questions about the reliability of studies that compare the coverage of GS, WoS, and Scopus, and conclude that GS collects significantly more citations [12,28,29] than its competitors. Further research should explore the citation counts of these databases to determine how comparable they are.

Title Errors

The main issue with the titles retrieved from GS is incompleteness, which causes problems such as false-positive matches. The similarity between the OHBM poster title and the JSUR article title demonstrates this GS difficulty. Other errors (typographical and spelling) cause problems in database management. More unwelcome is the missing title error. Instead of the title, 6.2% of references contained, for example, editor names, domain names, or ISBN numbers. These missing title errors raise several problems as follows: (1) references cannot be retrieved with a search by title, and (2) duplicates can be more frequent and more difficult to detect.

Author Errors

The citing documents examined were cowritten by 565 authors. Nevertheless, 124 (22.0%) authors were either incorrect or missing. These errors can cause problems in studies of the structure of scientific collaborative networks, which are commonly used graphs. Indeed, a fifth of the collaborative networks built may be incorrect and thus may generate imperfect relationships. First, the missing authors may truncate an important share of all the authors involved. Second, the irrelevant added authors may create a bias that a graph's algorithms can reinforce. Consequently, researchers may overestimate a relationship or ignore another determinant one.

Data Collection Errors

Duplicates, translations, and reprints are frequent in GS data. As collected data can be biased by duplicates, their detection is the first step implemented in studies using GS data. The duplicate, translation, and reprint rates found in this study were similar to those in previous studies [30,32]. In addition, the URL address of citing documents is commonly used to detect duplicates and collect full-text documents. Because a missing URL may cause difficulties, previous studies resolved this issue by automatically deleting a reference without a web address [16,32]. By contrast, in this study, 6.4% of URL addresses were missing, but only 0.7% of them could not be found with a manual search. Half of these found documents were usable, and half were duplicate, translation, and reprint references. Consequently, the duplicate search removed 7.8% of the references, whereas the irrelevant deletion of references without a URL address led to the omission of 3.2% of the citing documents. Again, this may cause biased results.

Publication Year Errors

Incorrect years had a lower frequency than missing years (3.3% and 8.0% of references, respectively). These missing values can cause major problems in data collection. As GS limits the search results to the first 1000 citing references per query, certain researchers have collected data by publication year to obtain a larger corpus of GS citations [25] or to focus their analysis on a specific period of time [45]. Other researchers have removed references containing incorrect publication years [16]. Thus, these neglected references may lead to truncated data and biased results. Inherent to the failed indexation process of GS, this publication year error may cause sampling errors that affect the representativeness of findings.

Data Verification Versus Biased Results

The GS error rate seems to be negligible when types of errors are considered in isolation. These types of claims have been made about false positives [20], duplicates [10], and incorrect publication years [2]. By contrast, with regard to GS errors, Harzing [55] argues fatalistically that "bibliometrics is an inexact science and that any data source has its own flaws." However, Hicks et al [40], in presenting the Leiden Manifesto,

XSL•FO
RenderX

emphasized the importance of the quality of the data used in research evaluation. Conversely, when the GS error rates are observed as a whole, a worrying cumulative effect is revealed. Indeed, only 2 of 281 (0.71%) references collected from GS were free from errors. This raises a question about the reliability of GS citation counts. In this study, 2 neuroimaging works were cited 281 times according to GS. However, this citation count is incorrect. In fact, these works were cited 131 times in academic publications (ie, excluding duplicates, reprints, translations, inaccessible documents, and false positives), which is 53.4% less than the GS claim. Thus, the full sample collected from GS (281 citations) can considerably differ from the proper sample (131 citations). There is thus a major risk of producing incorrect and biased results that do not accurately reflect the data examined.

Consequently, meticulous verification and cleaning of GS data are essential before using them. Considering this, several precautions should be taken to improve the reliability of GS data. First, detect and remove duplicate, translation, and reprint references and subsequently merge their citation counts. Second, consult the full-text documents of the full sample to remove false-positive matches. Third, verify the document type of each reference to exclude nonacademic publications.

Because results will be biased or wrong if these verification steps are not performed, is it possible to study a large-scale sample of GS citations (approximately several thousand)? It seems unlikely unless substantial resources are allocated for such verification. Indeed, Meho and Yang [21] were allowed 18 minutes per reference (3000 hours of work for 10,000 citation samples). In this study, a work time of 32 minutes per reference was necessary to complete the verification (150 hours for 281 citations). What about automatization of the verification? Studies that cleaned large-scale data either in part or as a whole using an automatic cleaning process [4,10,20] reported a lower error rate and fewer error types than studies using a manual cleaning process [30,54]. Therefore, it is reasonable to have doubts about the efficiency of this automatic cleaning.

Finally, studying a small sample of GS data seems more adequate than studying a large sample in terms of obtaining reliable data and accurate findings. Nevertheless, there is a need to conduct further research to develop statistical tools for weighting the correlation calculation in a large-scale sample of GS data, which are widely used in database coverage studies. However, these tools may not correct the collection issue inherent to the GS database.

Alternatively, according to the reference accuracy literature [9,10,12-15], databases, such as WoS and Scopus, can be used in place of GS because they are more reliable, though they have narrower coverage than GS. Indeed, WoS has an average error rate of 0.1% [4,18,31,32], and this rate is 1.0% for Scopus [10,31]. However, since GS is a free database [56], it may be the only possible way to conduct a study. However, knowing that all databases are likely to contain errors, verifying a sample of data is a useful precaution.

To conclude, the categorization of the errors encountered in the data extracted from GS provides researchers with methodological and epistemological reflections so that they become aware, with precision, of the probable errors that they are likely to encounter, and can consequently adjust their methodological choice. For example, the number of citations obtained by GS may not be completely accurate, or the names of the authors mentioned may not be completely correct. With a sample of several thousand references, these errors can have a noticeable impact on the results.

Conclusion

Almost all of the data retrieved from GS contained at least one error, calling the reliability of GS data into question. Further, the reliability of studies using a large-scale sample without verification and data cleaning is also called into question. Moreover, studies using GS to evaluate research activity or compare the coverage of several databases (ie, GS, WoS, and Scopus) may be affected by substantial biases, including citation miscounts.

However, researchers who are able to spend a considerable amount of time on the meticulous verification of their small samples can obtain various references for journal articles, books, edited book chapters, and conference proceedings from GS. This ability can be especially useful in bibliometric studies based on material published in research areas in which journal articles are less predominant than other publication types.

Limitations

Since the data used are limited and specific, the results obtained cannot be generalized. However, this case study provides a kind of “stress test” of GS to promote reflection on the limits of this free database.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Raw data.
[RAR File , 30 KB - jmir_v24i5e28354_app1.rar]

Multimedia Appendix 2
Number of errors detected per reference retrieved from Google Scholar.
[PDF File (Adobe PDF File), 94 KB - jmir_v24i5e28354_app2.pdf]

XSL•FO
RenderX

Multimedia Appendix 3
Data collection error: types and rates.
[PDF File (Adobe PDF File), 79 KB - jmir_v24i5e28354_app3.pdf]

Multimedia Appendix 4
Document types of references collected from Google Scholar as a function of 3 categories (academic journal, nonacademic journal, and Google Scholar non-"scholarly literature").
[PDF File (Adobe PDF File), 86 KB - jmir_v24i5e28354_app4.pdf]

Multimedia Appendix 5
Results of the Mann-Whitney test on the type of reference (academic publication or nonacademic document) and the number of Google Scholar errors, and data visualization with a box plot.
[PDF File (Adobe PDF File), 117 KB - jmir_v24i5e28354_app5.pdf]

Multimedia Appendix 6
Inaccurate content identified in the "Author" column retrieved from Google Scholar via Publish or Perish software.
[PDF File (Adobe PDF File), 97 KB - jmir_v24i5e28354_app6.pdf]

Multimedia Appendix 7
Inaccurate content identified in the "Title" column retrieved from Google Scholar via Publish or Perish software.
[PDF File (Adobe PDF File), 81 KB - jmir_v24i5e28354_app7.pdf]

Multimedia Appendix 8
Inaccurate content identified in the "Year" column retrieved from Google Scholar via Publish or Perish software.
[PDF File (Adobe PDF File), 80 KB - jmir_v24i5e28354_app8.pdf]

Multimedia Appendix 9
Content of the "Publication" column retrieved from Google Scholar via Publish or Perish software as a function of reference document type.
[PDF File (Adobe PDF File), 78 KB - jmir_v24i5e28354_app9.pdf]

Multimedia Appendix 10
Spelling and orthographical errors in the academic publications retrieved from Google Scholar via Publish or Perish software: journal name, edited book title, and conference proceedings book title.
[PDF File (Adobe PDF File), 100 KB - jmir_v24i5e28354_app10.pdf]

Multimedia Appendix 11
Content of the "Publisher" column retrieved from Google Scholar via Publish or Perish software as a function of reference document type.
[PDF File (Adobe PDF File), 77 KB - jmir_v24i5e28354_app11.pdf]

References

1. Orduna-Malea E, Martín-Martín A, Delgado López-Cózar E. Google Scholar as a source for scholarly evaluation: a bibliographic review of database errors. Revista Española de Documentación Científica 2017 Nov 22;40(4):185 [FREE Full text] [doi: 10.3989/redc.2017.4.1500]
2. Martin-Martin A, Orduna-Malea E, Harzing AW, Delgado López-Cózar E. Can we use Google Scholar to identify highly-cited documents? Journal of Informetrics 2017 Feb;11(1):152-163 [FREE Full text] [doi: 10.1016/j.joi.2016.11.008]
3. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. The FASEB Journal 2008 Feb;22(2):338-342 [FREE Full text] [doi: 10.1096/fj.07-9492LSF] [Medline: 17884971]
4. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015 Sep 17;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
5. Bornmann L, Marx W, Schier H, Rahm E, Thor A, Daniel HD. Convergent validity of bibliometric Google Scholar data in the field of chemistry—Citation counts for papers that were accepted by Angewandte Chemie International Edition or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts. Journal of Informetrics 2009 Jan;3(1):27-35 [FREE Full text] [doi: 10.1016/j.joi.2008.11.001]

XSL•FO
RenderX

6. Abad-García MF, González-Teruel A, González-Llinares J. Effectiveness of OpenAIRE, BASE, Recolecta, and Google Scholar at finding spanish articles in repositories. Journal of the Association for Information Science and Technology 2018 Apr;69(4):619-622 [FREE Full text] [doi: 10.1002/asi.23975]
7. Orduna-Malea E, Ayllón JM, Martín-Martín A, Delgado López-Cózar E. Methods for estimating the size of Google Scholar. Scientometrics 2015 May 31;104(3):931-949 [FREE Full text] [doi: 10.1007/s11192-015-1614-6]
8. Mingers J, Lipitakis EAECG. Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. Scientometrics 2010 Jul 11;85(2):613-625 [FREE Full text] [doi: 10.1007/s11192-010-0270-0]
9. Harzing AW, Alakangas S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics 2016 Feb;106(2):787-804 [FREE Full text] [doi: 10.1007/s11192-015-1798-9]
10. Moed HF, Bar-Ilan J, Halevi G. A new methodology for comparing Google Scholar and Scopus. Journal of Informetrics 2016 May;10(2):533-551 [FREE Full text] [doi: 10.1016/j.joi.2016.04.017]
11. Gehanno JF, Rollin L, Darmoni S. Is the coverage of Google Scholar enough to be used alone for systematic reviews. BMC Medical Informatics and Decision Making 2013 Dec;13(1):7 [FREE Full text] [doi: 10.1186/1472-6947-13-7] [Medline: 23302542]
12. Harzing AW. A preliminary test of Google Scholar as a source for citation data: a longitudinal study of Nobel prize winners. Scientometrics 2013 Mar;94(3):1057-1075 [FREE Full text] [doi: 10.1007/s11192-012-0777-7]
13. Mingers J, Xu F. The drivers of citations in management science journals. European Journal of Operational Research 2010 Sep;205(2):422-430 [FREE Full text] [doi: 10.1016/j.ejor.2009.12.008]
14. Franceschet M. A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. Scientometrics 2010 Apr;83(1):243-258 [FREE Full text] [doi: 10.1007/s11192-009-0021-2]
15. Harzing AW. A longitudinal study of Google Scholar coverage between 2012 and 2013. Scientometrics 2014 Jan;98(1):565-575 [FREE Full text] [doi: 10.1007/s11192-013-0975-y]
16. Prins AAM, Costas R, van Leeuwen TN, Wouters PF. Using Google Scholar in research evaluation of humanities and social science programs: A comparison with Web of Science data. Research Evaluation 2016 Jul;25(3):264-270 [FREE Full text] [doi: 10.1093/reseval/rvv049]
17. Kousha K, Thelwall M, Rezaie S. Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. Journal of the American Society for Information Science and Technology 2011 Nov;62(11):2147-2164 [FREE Full text] [doi: 10.1002/asi.21608]
18. de Winter JCF, Zadpoor AA, Dodou D. The expansion of Google Scholar versus Web of Science: a longitudinal study. Scientometrics 2014 Feb;98(2):1547-1565 [FREE Full text] [doi: 10.1007/s11192-013-1089-2]
19. Wildgaard L. A comparison of 17 author-level bibliometric indicators for researchers in Astronomy, Environmental Science, Philosophy and Public Health in Web of Science and Google Scholar. Scientometrics 2015 Sep;104(3):873-906 [FREE Full text] [doi: 10.1007/s11192-015-1608-4]
20. Franceschini F, Maisano D, Mastrogiacomo L. Empirical analysis and classification of database errors in Scopus and Web of Science. Journal of Informetrics 2016 Nov;10(4):933-953 [FREE Full text] [doi: 10.1016/j.joi.2016.07.003]
21. Meho LI, Yang K. Impact of data sources on citation counts and rankings of LIS faculty: Web of science versus scopus and google scholar. Journal of the American Society for Information Science and Technology 2007 Nov;58(13):2105-2125 [FREE Full text] [doi: 10.1002/asi.20677]
22. Li J, Sanderson M, Willett P, Norris M, Oppenheim C. Ranking of library and information science researchers: Comparison of data sources for correlating citation data, and expert judgments. Journal of Informetrics 2010 Oct;4(4):554-563 [FREE Full text] [doi: 10.1016/j.joi.2010.06.005]
23. Mingers J, O'Hanley JR, Okunola M. Using Google Scholar institutional level data to evaluate the quality of university research. Scientometrics 2017 Dec;113(3):1627-1643 [FREE Full text] [doi: 10.1007/s11192-017-2532-6] [Medline: 29200538]
24. Martín-Martín A, Orduna-Malea E, Delgado López-Cózar E. A novel method for depicting academic disciplines through Google Scholar Citations: The case of Bibliometrics. Scientometrics 2018 Mar;114(3):1251-1273 [FREE Full text] [doi: 10.1007/s11192-017-2587-4]
25. Martín-Martín A, Orduna-Malea E, Ayllón JM, Delgado López-Cózar E. A two-sided academic landscape: snapshot of highly-cited documents in Google Scholar (1950-2013). Rev. esp. doc. cient 2016 Dec 13;39(4):149 [FREE Full text] [doi: 10.3989/redc.2016.4.1405]
26. Pitol SP, De Groote SL. Google Scholar versions: do more versions of an article mean greater impact? Library Hi Tech 2014 Nov 11;32(4):594-611 [FREE Full text] [doi: 10.1108/LHT-05-2014-0039]
27. Jamali HR, Nabavi M. Open access and sources of full-text articles in Google Scholar in different subject fields. Scientometrics 2015 Dec;105(3):1635-1651 [FREE Full text] [doi: 10.1007/s11192-015-1642-2]
28. Lasda Bergman EM. Finding citations to social work literature: the relative benefits of using Web of Science, Scopus, or Google Scholar. The Journal of Academic Librarianship 2012 Nov;38(6):370-379 [FREE Full text] [doi: 10.1016/j.acalib.2012.08.002]

XSL•FO
RenderX

29. De Groote SL, Raszewski R. Coverage of Google Scholar, Scopus, and Web of Science: a case study of the h-index in nursing. Nursing Outlook 2012 Nov;60(6):391-400 [FREE Full text] [doi: 10.1016/j.outlook.2012.04.007] [Medline: 22748758]
30. Granter SR, Laga AC, Larson AR. Calciphylaxis and the persistence of medical misinformation in the era of Google. American Journal of Clinical Pathology 2015 Sep;144(3):427-431 [FREE Full text] [doi: 10.1309/AJCPDMWVGKW9N1CU] [Medline: 26276773]
31. Kulkarni AV, Aziz B, Shams I, Busse JW. Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. JAMA 2009 Sep 09;302(10):1092-1096 [FREE Full text] [doi: 10.1001/jama.2009.1307] [Medline: 19738094]
32. García-Pérez MA. Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. Journal of the American Society for Information Science and Technology 2010 Oct;61(10):2070-2085 [FREE Full text] [doi: 10.1002/asi.21372]
33. Bar-Ilan J. Which h-index? — A comparison of WoS, Scopus and Google Scholar. Scientometrics 2008 Jan;74(2):257-271 [FREE Full text] [doi: 10.1007/s11192-008-0216-y]
34. Valderrama-Zurián JC, Aguilar-Moya R, Melero-Fuentes D, Aleixandre-Benavent R. A systematic analysis of duplicate records in Scopus. Journal of Informetrics 2015 Jul;9(3):570-576 [FREE Full text] [doi: 10.1016/j.joi.2015.05.002]
35. Fiolet T, Guihur A, Rebeaud ME, Mulot M, Peiffer-Smadja N, Mahamat-Saleh Y. Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. Clinical Microbiology and Infection 2021 Jan;27(1):19-27 [FREE Full text] [doi: 10.1016/j.cmi.2020.08.022] [Medline: 32860962]
36. Oyibo K, Sahu KS, Oetomo A, Morita PP. Factors influencing the adoption of contact tracing applications: protocol for a systematic review. JMIR Research Protocols 2021 Jun 01;10(6):e28961 [FREE Full text] [doi: 10.2196/28961] [Medline: 33974551]
37. Oliveira J e Silva L, Maldonado G, Brigham T, Mullan AF, Utengen A, Cabrera D. Evaluating scholars' impact and influence: Cross-sectional study of the correlation between a novel social media-based score and an author-level citation metric. Journal of Medical Internet Research 2021 May 31;23(5):e28859 [FREE Full text] [doi: 10.2196/28859] [Medline: 34057413]
38. Cole S. Citations and the evaluation of individual scientists. Trends in Biochemical Sciences 1989 Jan;14(1):9-13 [FREE Full text] [doi: 10.1016/0968-0004(89)90078-9]
39. Sauvayre R. De la diffusion scientifique aux normes de scientificité ("The scientific dissemination of scientific norms"). Habilitation à diriger des recherches ("Accreditation to Supervise Research (HDR) dissertation"). Paris, France: Sorbonne University; 2020 Dec. URL: http://romy.sauvayre.online.fr/Resume_20_HDR.en.php [accessed 2022-05-10]
40. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for research metrics. Nature 2015 Apr 23;520(7548):429-431 [FREE Full text] [doi: 10.1038/520429a] [Medline: 25903611]
41. Bennett CM, Baird AA, Miller MB, Wolford GL. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction. 2009 Presented at: 15th Annual Meeting of the Organization for Human Brain Mapping; June 18-23, 2009; San Francisco, CA URL: http://prefrontal.org/files/posters/Bennett-Salmon-2009.pdf
42. Bennett CM, Baird AA, Miller MB, Wolford GL. Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: an argument for proper multiple comparisons correction. Journal of Serendipitous and Unexpected Results 2010;1(1):1-5 [FREE Full text]
43. Inclusion Guidelines for Webmasters. Google Scholar. URL: https://scholar.google.com/intl/en/scholar/inclusion.html#content [accessed 2018-01-10]
44. Harzing AW. Accuracy: PoP vs GS for journals. Harzing.com. Research in International Management. URL: https://harzing.com/resources/publish-or-perish/tutorial/accuracy/pop-vs-gs-for-journals [accessed 2018-02-15]
45. Thelwall M, Kousha K. ResearchGate versus Google Scholar: Which finds more early citations? Scientometrics 2017 Aug;112(2):1125-1131 [FREE Full text] [doi: 10.1007/s11192-017-2400-4]
46. Adams D. Google Scholar. Harzing.com. Research in International Management. 2016 Dec 20. URL: https://harzing.com/resources/publish-or-perish/manual/using/data-sources/google-scholar [accessed 2019-03-19]
47. Cole S, Cole JR. Scientific output and recognition: a study in the operation of the reward system in science. American Sociological Review 1967 Jun;32(3):377-390 [FREE Full text] [doi: 10.2307/2091085] [Medline: 6046811]
48. de Solla Price DJ. Networks of scientific papers. Science 1965 Jul 30;149(3683):510-515 [FREE Full text] [doi: 10.1126/science.149.3683.510] [Medline: 14325149]
49. About Google Scholar. Google Scholar. URL: https://scholar.google.com/intl/en/scholar/about.html [accessed 2018-01-10]
50. Google Scholar Search Tips. Google Scholar. URL: https://scholar.google.com/intl/en/scholar/help.html#coverage [accessed 2021-07-12]
51. Adams D. Results list. Harzing.com. Research in International Management. 2016 Dec 16. URL: https://harzing.com/resources/publish-or-perish/manual/using/query-results/results-list [accessed 2018-02-15]
52. Guagliardo P, Libkin L. Correctness of SQL Queries on databases with nulls. ACM SIGMOD Record 2017 Oct 31;46(3):5-16 [FREE Full text] [doi: 10.1145/3156655.3156657]

XSL•FO
RenderX

53. Baneyx A. "Publish or Perish" as citation metrics used to analyze scientific output in the humanities: International case studies in economics, geography, social sciences, philosophy, and history. Archivum Immunologiae et Therapiae Experimentalis 2008 Dec;56(6):363-371 [FREE Full text] [doi: 10.1007/s00005-008-0043-0] [Medline: 19043670]
54. Bar-Ilan J. Citations to the “Introduction to informetrics” indexed by WOS, Scopus and Google Scholar. Scientometrics 2010 Mar;82(3):495-506 [FREE Full text] [doi: 10.1007/s11192-010-0185-9]
55. Harzing AW. Accuracy. Harzing.com. Research in International Management. 2016 Dec 20. URL: https://harzing.com/resources/publish-or-perish/manual/using/query-results/accuracy [accessed 2018-02-15]
56. Basu A, Malhotra D, Seth T, Muhuri PK. Global distribution of Google Scholar citations: a size-independent institution-based analysis. Journal of Scientometric Research 2019 Sep 03;8(2):72-78 [FREE Full text] [doi: 10.5530/jscires.8.2.12]

Abbreviations

GS: Google Scholar
JSUR: Journal of Serendipitous and Unexpected Results
OHBM: Organization for Human Brain Mapping
PoP: Publish or Perish software
WoS: Web of Science

Edited by A Mavragani; submitted 02.03.21; peer-reviewed by Z Ren, T Capote, S Watty, S Wei, A Roy, A Louren, J Marconato; comments to author 19.04.21; revised version received 16.07.21; accepted 15.03.22; published 27.05.22.

Please cite as:
Sauvayre R
Types of Errors Hiding in Google Scholar Data
J Med Internet Res 2022;24(5):e28354
URL: https://www.jmir.org/2022/5/e28354
doi:10.2196/28354
PMID:35622395

©Romy Sauvayre. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2022. 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.

XSL•FO
RenderX

Original Paper

Delivering Mindfulness-Based Interventions for Insomnia, Pain, and Dysfunctional Eating Through a Text Messaging App: Three Randomized Controlled Trials Investigating the Effectiveness and Mediating Mechanisms

Amanda CM Li[1], MSSc; Keith KL Wong[2], MSSc; Floria HN Chio[3], PhD; Winnie WS Mak[1], PhD; Loretta WH Poon[4], MA

[1]Department of Psychology, The Chinese University of Hong Kong, Hong Kong, Hong Kong

[2]Clinical Psychological Services, New Life Psychiatric Rehabilitation Association, Hong Kong, Hong Kong

[3]Department of Counselling and Psychology, Hong Kong Shue Yan University, Hong Kong, Hong Kong

[4]newlife.330, New Life Psychiatric Rehabilitation Association, Hong Kong, Hong Kong

Corresponding Author:
Winnie WS Mak, PhD
Department of Psychology
The Chinese University of Hong Kong
Shatin, NT
Hong Kong
Hong Kong
Phone: 852 39436577
Fax: 852 26035019
Email: wwsmak@psy.cuhk.edu.hk

Abstract

Background: Although text messaging has the potential to be the core intervention modality, it is often used as an adjunct only. To improve health and alleviate the distress related to insomnia, pain, and dysregulated eating of people living in urban areas, text messaging–based mindfulness-based interventions were designed and evaluated in 3 randomized controlled trials.

Objective: This study investigated the effectiveness and mediating mechanisms of text messaging–based mindfulness-based interventions for people with distress related to insomnia, pain, or dysregulated eating.

Methods: In these trials, 333, 235, and 351 participants were recruited online and randomized to intervention and wait-list control conditions for insomnia, pain, and dysregulated eating, respectively. Participants experienced 21 days of intervention through WhatsApp Messenger. Participants completed pre-, post-, 1-month follow-up, and 3-month follow-up self-report questionnaires online. The retention rates at postmeasurements were 83.2% (139/167), 77.1% (91/118), and 72.9% (129/177) for intervention groups of insomnia, pain, and dysregulated eating, respectively. Participants' queries were answered by a study technician. Primary outcomes included insomnia severity, presleep arousal, pain intensity, pain acceptance, and eating behaviors. Secondary outcomes included mindfulness, depression, anxiety, mental well-being, and functional impairments. Mindfulness, dysfunctional beliefs and attitudes about sleep, pain catastrophizing, and reactivity to food cues were hypothesized to mediate the relationship between the intervention and outcomes.

Results: For all 3 studies, the intervention groups showed significant improvement on most outcomes at 1-month follow-up compared to their respective wait-list control groups; some primary outcomes (eg, insomnia, pain, dysregulated eating indicators) and secondary outcomes (eg, depression, anxiety symptoms) were sustained at 3-month follow-up. Medium-to-large effect sizes were found at postassessments in most outcomes in all studies. In the intervention for insomnia, mediation analyses showed that dysfunctional beliefs and attitudes about sleep mediated the effect of the intervention on all primary outcomes and most secondary outcomes at both 1-month and 3-month follow-ups, whereas mindfulness mediated the intervention effect on presleep arousal at 1-month and 3-month follow-ups. In the intervention for pain, pain catastrophizing mediated the effect of intervention on pain intensity and functioning at both 1-month and 3-month follow-ups, whereas mindfulness only mediated the effect of intervention on anxiety and depressive symptoms. In the intervention for dysregulated eating, power of food mediated the effect of intervention

XSL•FO
RenderX

on both uncontrolled and emotional eating at both 1-month and 3-month follow-ups and mindfulness was found to mediate the effect on depressive symptoms at both 1-month and 3-month follow-ups.

Conclusions: These 3 studies converged and provided empirical evidence that mindfulness-based interventions delivered through text messaging are effective in improving distress related to sleep, pain, and dysregulated eating. Text messaging has the potential to be a core intervention modality to improve various common health outcomes for people living a fast-paced lifestyle.

Trial Registration: Clinical Research and Biostatistics Clinical Trials Registry CUHK_CCRB00559; https://tinyurl.com/24rkwarz

(J Med Internet Res 2022;24(5):e30073) doi:10.2196/30073

KEYWORDS

text messaging; mindfulness; insomnia; pain; dysregulated eating; mHealth; mental health; SMS; distress; intervention; outcome; mobile interventions

Introduction

Prevalence of Insomnia, Pain, and Dysregulated Eating

People living in urban areas have prevailing complaints of stress and related health concerns. For instance, the general population of Hong Kong has a stressful life, with an estimation of 13.3% having common mental disorders [1]. Stress was found to be correlated with multiple health issues, including insomnia, chronic pain, and unhealthy eating behaviors [2-4]. In the United Kingdom, the economic burden of low back pain, insomnia, and eating-related conditions was estimated to be £2.79 billion, £46.3 billion, and £8.5 billion, respectively (US $1=£0.77), projected into 2018 costs [5-7]. In Hong Kong, insomnia and chronic pain affect 35.2% and 39.4% of the general population, respectively, and close to half (46.9%) of the general population was found to have unhealthy eating habits [8-10].

Mindfulness-Based Interventions for Insomnia, Pain, and Dysregulated Eating

To tackle these health concerns, mindfulness-based interventions (MBIs) have been applied to alleviate stress and enhance well-being. Mindfulness is defined as "paying attention in a particular way on purpose in the present moment and nonjudgmentally" [11]. In particular, MBIs were found to be effective in improving insomnia [12,13], pain [14-16], binge eating, and emotional eating [17,18]. MBIs have not only been applied to clinical populations, but they have also been demonstrated to be beneficial to the well-being of nonclinical populations. Meta-analyses showed that MBIs have moderate effect size in reducing stress, psychological distress, depression, and anxiety among healthy individuals [19,20]. Internet-based MBIs are gaining more evidence as well. According to another review and meta-analysis, web-based MBIs are effective in reducing depressive symptoms, anxiety symptoms, stress, and improving well-being and mindfulness with small-to-medium effect size [21]. The above evidence provides the foundation to further develop internet-based MBIs for both clinical and healthy individuals.

Mechanism of MBI

Regarding the mechanisms of MBIs in the promotion of well-being and the reduction of distress, the cultivation of mindfulness is found to mediate the relationship between MBIs and various outcome variables. For instance, Nyklíček and Kuijpers [22] found that changes in mindfulness partially mediated the relationship between mindfulness-based stress reduction intervention and its positive effects among people with distress symptoms. In addition to changes in mindfulness, according to Shapiro et al [23], the cultivation of mindfulness may also facilitate "reperceiving," which is a shift of perspective that leads to an increased capacity for relating to one's internal or external experiences objectively. For insomnia, mindfulness allows people to respond to stressors more skillfully by changing the patterns of worry and rumination that improve sleep quality [24]. For pain, preliminary evidence suggested that mindfulness and pain catastrophizing mediated the relationship between interventions and reduced perceived stress and improved quality of life [22,25]. For dysfunctional eating, MBI regulates appetitive and emotional processes by increasing both awareness and sensitivity to the eating process such that people can disengage themselves from the reactive eating habits [26]. To further test these mechanisms in this study, we hypothesized that text messaging–based MBIs would promote better well-being and fewer symptoms of insomnia, pain, and unhealthy eating through changes in both mindfulness and reperceiving.

Text Messaging As the Core of Treatment Modality

Although internet-based MBIs were effective in reducing distress and promoting well-being [21], users' retention and engagement remain challenging for digital health interventions [27,28]. A median of 56% retention was found, and the attrition rate can be as high as 75% for internet-based interventions [29,30]. Text messaging, considered as one of the future trends of internet-based interventions, may alleviate this problem, given the temporal synchronization feature of instant messaging that allows the intervention to catch people's attention directly [31,32]. A systematic review showed that text messaging improves treatment adherence and reduces social isolation [33]. Up till now, to the best of our knowledge, text messaging is mostly used as a reminder, self-monitoring tool, or as an adjunct to face-to-face intervention instead of being utilized as the core treatment modality, except in a few physical health intervention studies [29,34-38]. Thus, the potential for text messaging to offer mental health intervention that improves users' engagement and retention is yet to be investigated.

Aim of This Study

The aim of this study was to test whether a text messaging–based MBI is effective in reducing distress related to insomnia, pain, and dysfunctional eating in 3 randomized controlled trials.

XSL•FO
RenderX

WhatsApp [39] was chosen as the messaging tool because it is the most popular in the world with 2000 million monthly active users and is the most accessible option for the community [40]. Specifically, we aimed to investigate whether text messaging–based MBI can result in improvement in primary outcomes (ie, insomnia severity, pain severity, dysregulated eating) and secondary outcomes (ie, depressive and anxiety symptoms, mental well-being, and functional impairment). We also aimed to examine mindfulness, dysfunctional beliefs (for insomnia), pain catastrophizing (for chronic pain), and power of food (for dysregulated eating) as possible mechanisms of change that mediate the relationship between text messaging–based MBI with both primary and secondary outcomes.

Methods

Procedure

WhatsApp numbers for each trial were disseminated in the recruitment materials, and interested participants enrolled in the relevant trial via the instant messenger app, that is, WhatsApp. The nature and procedure of the studies were then explained to the participants through a WhatsApp message. Participants then completed the self-assessment web-based questionnaire together with a written consent page. Screening for eligibility of the participants was done upon completion of the preintervention questionnaire. Eligible participants were grouped into monthly batches, and the second author handled the enrollment, randomization, and intervention assignment of the participants. The randomization was conducted with Microsoft Office Excel 2010 [41] on an individual basis and 1:1 ratio between the intervention or wait-list control conditions with stratification on age (cutoff at 45 years old) and prior experience of mindfulness practice (no experience at all vs not novice). Participants were informed about their allocation of condition via WhatsApp by the same person who did the randomization. No blinding was feasible given the use of wait-list control in the design. A feasibility study on text messaging–based MBI had been done prior to this study. The intervention had a similar structure with this study. It was an uncontrolled study aiming at examining the impact of a general MBI, and the results revealed an improvement in the well-being and mindfulness of the participants. Therefore, the research team proceeded to these randomized controlled trial studies. Preintervention, postintervention, 1-month, and 3-month follow-up questionnaires were distributed via WhatsApp using Qualtrics [42]. Pretesting of the questionnaire before launch was done by the second author. The questionnaires were divided into 4 pages, namely, the consent form, demographic and screening items, primary outcomes, and secondary outcomes. The questionnaire pages were limited to around 50 items per page. Cookies were enabled, and participants could resume their questionnaire on the same device. Participants were required to input their research ID in the questionnaire, and duplicated data with the same research ID were eliminated. The postintervention questionnaires were sent to participants in both wait-list control and intervention conditions on the 22nd day. To adopt an intent-to-treat approach, all participants were invited to complete the self-report questionnaires at every time point regardless of whether they had completed the questionnaires at previous time points. Wait-list control participants received the intervention after completion of the 3-month follow-up questionnaire. Data were exported and stored as a password-protected Excel file in an encrypted flash drive.

Ethics Approval

These studies were conducted in accordance with the ethical codes of the Declaration of Helsinki. Ethics approval was obtained from the Joint Chinese University of Hong Kong-New Territories East Cluster Clinical Research Ethics Committee (2017.203-T). Findings in this paper are reported in accordance with the recommendations of the Consolidated Standards of Reporting Trials (CONSORT).

Participants

Participants were recruited from the general public via social networking platforms, a health-related information website, a local magazine, mass emails, and announcement postings at local tertiary institutions, collaborating nongovernmental organizations, and mutual-aid groups. Participants were quasi-anonymous. Multiple registrations to different trials were eliminated. Participants were only allowed to join one of the trials. The recruitment and follow-up lasted from July 2017 to October 2018; HKD 100 (US $12.87) was offered to 10 participants from each trial as an incentive via random draw. Eligibility criteria for participation included (1) age of 18 years or older, (2) ability to understand Cantonese and give consent, and (3) adequate level of computer literacy to follow the web-based instructions independently, together with daily access to the internet. Participants who self-reported receiving psychiatric services or active suicidality were excluded from this study; hotline and related resources were provided to those participants. No harm or other unintended effects were noticed in both conditions across all 3 trials. In the 3 trials, 364, 264, and 371 eligible participants were recruited for insomnia, pain, and dysregulated eating trials, respectively. Some of the participants withdrew from the studies or became out of reach after the assignment of condition. See Figure 1, Figure 2, and Figure 3 for the CONSORT diagrams of the 3 randomized controlled trials.

XSL•FO
RenderX

Figure 1. CONSORT diagram: enrollment and flowchart of study 1.

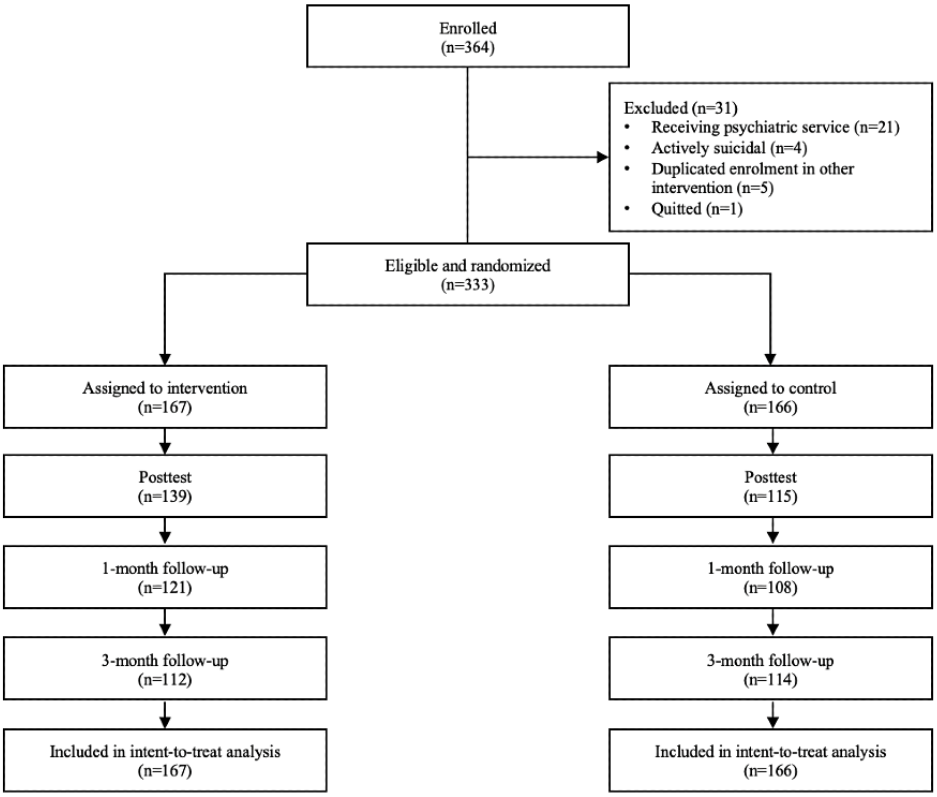
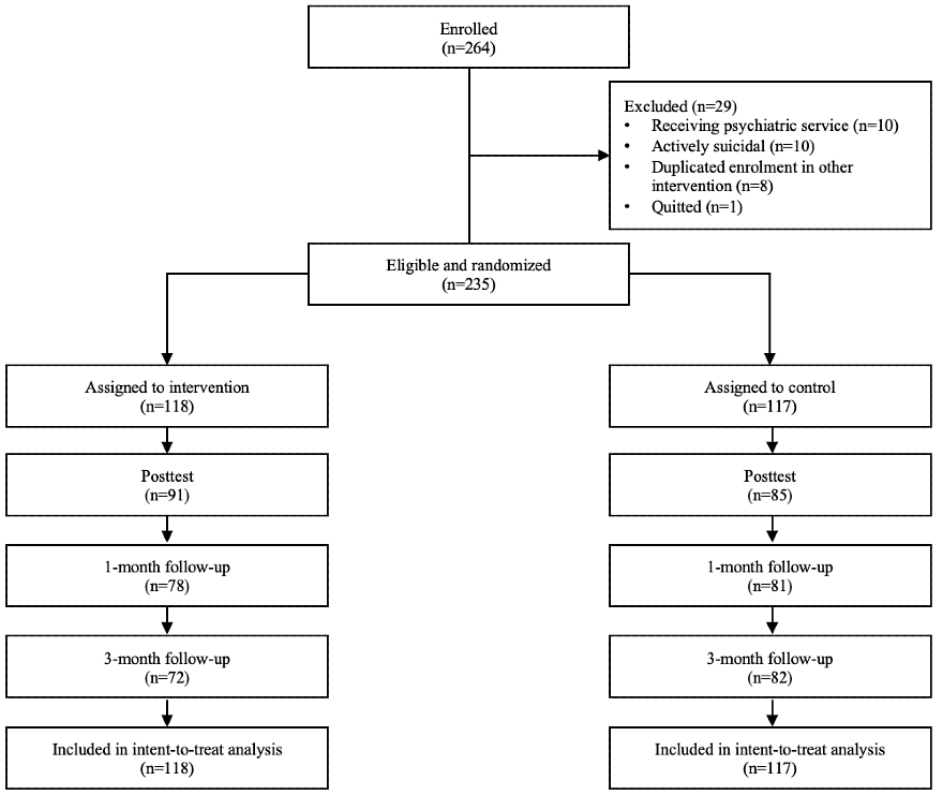
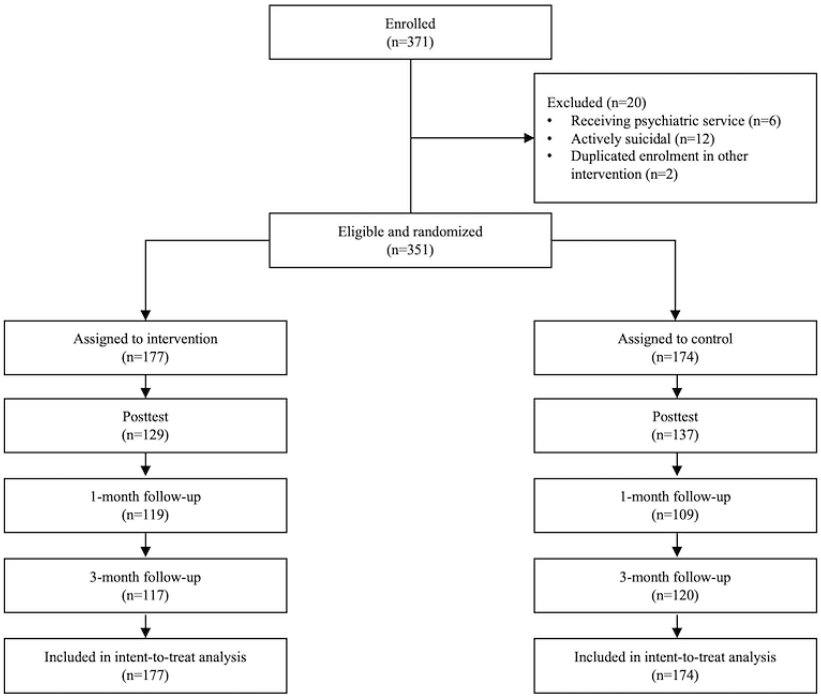


Figure 2. CONSORT diagram: enrollment and flowchart of study 2.



XSL•FO
RenderX

Figure 3. CONSORT diagram: enrollment and flowchart of study 3.



Intervention

All MBIs were delivered in 3-week packages, with 21 days of daily mindfulness exercises that consist of an audio clip and a psychoeducational article delivered via WhatsApp. Participants were grouped into monthly batches, and they received the same messages on a fixed schedule. The intervention was technician-assisted when participants had any specific questions. One technician was involved, and the interaction was confined to confirmation of enrollment, explanation of the study procedures, general enquiry, and reminder for data collection as well as monitoring adherence. Standardized materials were delivered with broadcast function without other personalized messages. To enhance adherence, brief mindfulness exercises were used, which took around 10-15 minutes to complete. Participants were invited to indicate their adherence by replying on WhatsApp after finishing each exercise. Reminder messages were sent to those participants who had not completed any exercise within a week. See Multimedia Appendix 1 and Multimedia Appendix 2 for the outlines of the interventions. The delivery format referenced a WhatsApp-based intervention of Acceptance and Commitment Therapy with daily passage and audio clip [43]. Participants received text and audio from the research team through WhatsApp. Notifications were pushed. Progress could be tracked by checking if the participants had seen the message and listened to the audio in WhatsApp. See Multimedia Appendix 3 for the screenshots.

For insomnia and pain, the intervention content was designed by clinical psychologists referencing the mindfulness-based stress reduction program [44]. Initially, the intervention cultivates participants' focused observation and awareness on bodily sensations, emotions, as well as thoughts. Practices included mindful breathing, mindful stretching, and body scan. Examples that are specific to insomnia and pain were included to increase relevance. Participants were encouraged to stay with aversive experiences with an open attitude. Lastly, suggestions were introduced to participants for integrating mindfulness practices in their daily life. For dysregulated eating, the MBI for eating incorporated 2 components running in parallel: (1) general mindfulness exercises and (2) specific eating-related components drawn from the Mindfulness-based Eating Awareness Training [26]. General mindfulness exercises started with observation and awareness of bodily sensations, emotions, and thoughts. Exercises on responding to unpleasant experiences were introduced. Meanwhile, eating-specific exercises were introduced to facilitate awareness of the eating experience. Specific exercises covered physical sensations of hunger and fullness, satisfaction of taste, changes in overall physical status, emotional triggers of eating, as well as choosing food with mindful awareness.

Measurements

Three sets of questionnaires were designed for each trial. Primary outcomes were used only in specific trials that included the severity of the targeted health concern itself and the attitude toward the health concern. Secondary outcomes were used across 3 trials that covered functional impairment, emotional disturbance, and mental well-being, while mindfulness and specific reperceiving variables were measured as potential mediators. Participants were also invited to provide demographic information such as age, gender, religious belief, education level, employment status, personal income, and prior experience in meditation. Usage of medication to cope with sleep disturbance or pain was also measured. Use of mental health services was also recorded.

Primary Outcomes of Study 1 on Insomnia

Insomnia Severity

Severity of insomnia was measured with the Insomnia Severity Index [45]. The Insomnia Severity Index consists of 7 items on

XSL•FO
RenderX

a 5-point Likert scale. It is widely used as a screening tool and outcome measure [46-48]. The scale covers difficulties in falling asleep or maintaining sleep, early wakening, satisfaction with sleep quality, daytime functioning, and quality of life. The items constitute a single overall score, with higher scores indicating greater severity of the sleep problem. A locally validated translated Chinese version was used in this study [49]. The Chinese version of the Insomnia Severity Index demonstrated satisfactory internal consistency (Cronbach α=.83), 2-week test-retest reliability (r=0.79), and concurrent validity [50]. In this study, the scale achieved good internal consistency (Cronbach α=.80).

Presleep Arousal

The Presleep Arousal Scale [51] is a 16-item 5-point Likert scale that consists of 2 subscales measuring somatic and cognitive arousal when getting to sleep. Higher scores indicate higher arousal while getting to sleep. In this study, the scale was translated into Chinese with Brislin's [52] forward and backward translation method. The 2 subscales of the Presleep Arousal Scale demonstrated satisfactory internal consistency in both the original validation study (Cronbach α=.76 and .81) [51] and this study (Cronbach α=.80 and .90).

Dysfunctional Beliefs and Attitude About Sleep (Reperceiving Mediator)

The Dysfunctional Beliefs and Attitude about Sleep Scale [53] is a 16-item scale measuring maladaptive beliefs and emotions about sleep difficulties. An 11-point Likert scale (0 "strongly disagree" to 10 "strongly agree") is used, and higher scores indicate more dysfunctional thoughts and emotions. A validated 16-item abbreviated Taiwanese Chinese version of the Dysfunctional Beliefs and Attitude about Sleep [54] scale demonstrated satisfactory internal consistency in both the Taiwan validation study (Cronbach α=.87) and this study (Cronbach α=.82).

Primary Outcomes of Study 2 on Pain

Pain Intensity

Subjective pain intensity was measured with a single-item Visual Analogue Scale (VAS). Participants were asked to indicate their overall pain intensity in the last week on a sliding scale from 0 to 100, with the anchor of "no pain," "moderate pain," and "extreme pain" on 0, 50, and 100, respectively. VAS is the most widely used tool for pain, and studies showed that VAS is valid and reliable tool that is sensitive in detection [55-57].

Pain Acceptance

The Chronic Pain Acceptance Questionnaire [58] was used to measure the willingness to accept pain and to maintain daily engagement despite the pain. This scale consists of 20 items on a 7-point Likert scale with 2 subscales: (1) activity engagement and (2) pain willingness. The overall score is based on the sum of the 2 subscale scores, with higher scores of Chronic Pain Acceptance Questionnaire indicating better acceptance of pain. The Chinese version of Chronic Pain Acceptance Questionnaire has been validated with good test-retest reliability (r=0.79), internal consistency (Cronbach α=.79), and concurrent validity [59]. In this study, the scale demonstrated good overall internal consistency (Cronbach α=.83).

Pain Catastrophizing (Reperceiving Mediator)

The Pain Catastrophizing Scale [60] was used to measure catastrophic beliefs and appraisals about pain. It consists of 13 items on a 5-point Likert scale, with higher scores indicating worse catastrophizing tendency toward pain. In this study, a locally validated Chinese version of the scale was used [61]. The scale demonstrated good internal consistency in both the validation study (Cronbach α=.93) and this study (Cronbach α=.94).

Primary Outcomes of Study 3 on Dysregulated Eating

Eating Behaviors

Participants' eating behaviors were measured by the 18-item Three-Factor Eating Questionnaire-Revised (TFEQ-R18) [62]. It consists of 18 four-point Likert scale items. The TFEQ-R18 consists of 3 subscales: (1) uncontrolled eating that refers to difficulties in regulating eating behaviors, (2) cognitive restraint that refers to conscious effort to inhibit food intake, and (3) emotional eating that refers to eating behaviors motivated by dysphoric mood, loneliness, or anxiety. In this study, the TFEQ-R18 has been translated into Chinese with Brislin's [52] translation and back translation method. The scales demonstrated adequate-to-good internal consistency (Cronbach α=.76-.85) in the original validation study [62]. In this study, the scale of uncontrolled eating and emotional eating demonstrated good internal consistency (Cronbach α=.83-.88). However, the scale of cognitive restraint demonstrated acceptable internal consistency (Cronbach α=.68).

Reactivity to Food Cue (Reperceiving Mediator)

The Power of Food Scale [63] was used to measure food craving across contexts with different levels of proximity to food. It consists of 15 items on a 5-point Likert scale. Research results supported a single-factor model [64], and higher scores indicate stronger reactivity to food cue. The literature revealed satisfactory internal consistency of the scale (Cronbach α=.81-.91). In this study, the scale has been translated into Chinese using Brislin's [52] forward and backward translation procedures. The translated scale demonstrated good internal consistency in this study (Cronbach α=.92).

Secondary Outcomes in All Studies

Mindfulness (Mediator)

The Mindful Attention Awareness Scale [65] is a 15-item scale that measures trait mindfulness characterized by a present-oriented attention and awareness. All items are rated on a 6-point Likert scale, and higher scores indicate higher mindful awareness. The scale has been validated for use with healthy normal adults [65]. Research evidence supported a single-factor solution, and the scale demonstrated good internal consistency, test-retest reliability, discriminant, convergent validity, and criterion validity. It could also differentiate meditators from nonmeditators [65,66]. In this study, a Taiwanese Chinese version of the scale [67] was used that

XSL•FO
RenderX

demonstrates good reliability in the Taiwan study (Cronbach α=.87) and this study (Cronbach α=.91).

Depression

The Patient Health Questionnaire [68] is a 9-item screening instrument for depressive symptoms. All items are rated on a 4-point Likert scale, and higher scores indicate more severe depressive symptoms. An officially translated Hong Kong traditional Chinese version was used in this study. Literature reported good internal consistency (Cronbach α=.86-.89) and criterion validity with 88% sensitivity and 88% specificity. In this study, the scale demonstrated good internal consistency (Cronbach α=.85).

Anxiety

The Generalized Anxiety Disorder scale [69] is a 7-item screening questionnaire for general anxiety symptoms. All items are rated on a 4-point Likert scale, with higher scores indicating more severe anxiety symptoms. An officially translated Hong Kong traditional Chinese version was used in this study. The scale demonstrated good internal consistency in both the original validation study (Cronbach α=.92) and this study (Cronbach α=.926). Literature reported good test-retest reliability (r=0.83) and criterion validity with 88% sensitivity and 89% specificity [69].

Mental Well-being

The World Health Organization Well-being Index [70] is a 5-item scale measuring mental well-being. All items are rated on a 6-point Likert scale and constitute a sum score. Higher scores indicate better mental well-being. An officially translated Chinese version was used in this study. Literature reported good internal consistency, concurrent validity, and sensitivity to change of the scale [71-73]. In this study, the scale demonstrated satisfactory internal consistency (Cronbach α=.91).

Functional Impairment

The Work and Social Adjustment Scale [74] measures functional impact due to health-related issues. The scale was used in the insomnia and pain studies. This scale consists of 5 items on different domains of functioning such as work, social, home, and leisure. All items were rated on a 9-point Likert scale, and higher scores indicate more severe impairment. Research evidence supported a single-factor solution of the scale and reported good internal consistency (Cronbach α=.93) [75]. A Chinese version of the scale was used in this study, which demonstrated good internal consistency (Cronbach α=.88).

Analysis

Significance tests such as the independent 2-sided t tests and chi-square test of independence were employed to explore any baseline difference between completers and dropout participants. Intent-to-treat analyses were used in this study. Data of all the participants were included in the analysis regardless of their treatment adherence or attrition. Missing data were treated using multiple imputations, and 100 imputed data sets were generated. van Ginkel and Kroonenberg's [76] method of repeated measures analysis of variance (ANOVA) using imputed data was used in this study. Specifically, missing values were estimated and 100 plausible complete versions of the data sets were created. Results from these 100 data sets were then pooled into 1 analysis by applying Rubin's [77] pooling procedures. The pooling procedure was carried out using an SPSS macro by van Ginkel [78]. Treatment condition (intervention condition vs wait-list control) and time were entered as the fixed between-group and within-subject factors, respectively. To examine the mediating effects of mindfulness and reperceiving (ie, dysfunctional beliefs and attitudes about sleep, pain catastrophizing, power of food) on the relationship between condition and the primary and secondary outcomes, path analyses were conducted using Mplus 7 [79]. In the path analyses, condition was dummy coded (with the wait-list control condition coded as the reference group) and was treated as the independent variable. Mindfulness and condition-specific reperceiving mediator at postintervention assessment were treated as the mediators. Primary and secondary outcomes at 1-month and 3-month follow-up assessments were treated as the dependent variables. Baseline scores of all variables included in the model were controlled.

Results

Participant Characteristics

The demographic data and baseline characteristics of the participants are summarized in Table 1. Overall, the dropout rate was consistent between the intervention and wait-list control conditions (insomnia: χ^2_1 (N=333)=0.2; P=.62; pain: χ^2_1 (N=235)=1.2; P=.27; dysregulated eating: χ^2_1 (N=351)=0.1; P=.80). Chi-square tests and independent t tests revealed no significant association between the demographics variable and the dropout rate. No significant baseline difference between completers and dropout participants on any outcome measures was found.

XSL•FO
RenderX

Table 1. Demographic data and baseline characteristics of the participants.

Characteristics	Study 1: Insomnia (N=333)		Study 2: Pain (N=235)		Study 3: Dysregulated eating (N=351)	
	Intervention	Wait-list control	Intervention	Wait-list control	Intervention	Wait-list control
Age (years), mean (SD)	41.67 (13.57)	42.59 (13.08)	41.19 (15.45)	41.51 (13.66)	36.32 (11.64)	35.69 (11.55)
Sex, n (%)						
Male	40 (24)	27 (16.2)	17 (14.4)	21 (18)	20 (11.3)	16 (9.2)
Female	126 (75.5)	139 (83.7)	101 (85.6)	96 (82.1)	157 (88.7)	158 (90.8)
Religion, n (%)						
Catholic	7 (4.2)	16 (9.6)	11 (9.3)	7 (6)	5 (2.8)	5 (2.9)
Christian	41 (24.6)	29 (17.5)	29 (24.6)	29 (24.8)	32 (18.1)	42 (24.1)
Buddhist	25 (15)	37 (22.3)	18 (15.3)	19 (16.2)	29 (16.4)	28 (16.1)
Others	1 (0.6)	1 (0.6)	0 (0)	0 (0)	1 (0.6)	2 (1.2)
None	92 (55.1)	83 (50)	60 (50.9)	62 (53)	110 (62.2)	97 (55.8)
Education, n (%)						
Primary	1 (0.6)	2 (1.2)	3 (2.5)	4 (3.4)	1 (0.6)	0 (0)
Junior secondary	7 (4.2)	6 (3.6)	4 (3.4)	8 (6.8)	2 (1.1)	0 (0)
Senior secondary	36 (21.6)	41 (24.7)	30 (25.4)	21 (18)	37 (20.9)	30 (17.2)
College/university	77 (46.1)	72 (43.4)	48 (40.7)	44 (37.6)	91 (51.4)	91 (52.3)
Master or above	43 (25.8)	43 (25.9)	31 (26.3)	40 (34.2)	46 (26)	53 (30.5)
Others	2 (1.2)	2 (1.2)	2 (1.7)	0 (0)	1 (0.6)	0 (0)
Employment, n (%)						
Student	18 (10.8)	15 (9)	7 (5.9)	10 (8.6)	31 (17.5)	30 (17.2)
Full-time	105 (62.9)	94 (56.6)	66 (55.9)	63 (53.9)	110 (62.2)	111 (63.8)
Part-time	8 (4.8)	9 (5.4)	10 (8.5)	11 (9.4)	13 (7.3)	13 (7.5)
Unemployed	3 (1.8)	1 (0.6)	4 (3.4)	3 (2.6)	0 (0)	4 (2.3)
Retired	19 (11.4)	25 (15.1)	19 (16.1)	14 (12)	5 (2.8)	7 (4)
Others	9 (5.4)	16 (9.6)	11 (9.3)	12 (10.3)	15 (8.5)	7 (4)
Previous experience in mindfulness practice, n (%)						
Yes	40 (24)	42 (25.3)	33 (28)	32 (27.4)	41 (23.2)	43 (24.7)
No	126 (75.5)	124 (74.7)	85 (72)	85 (72.6)	136 (76.8)	131 (75.3)

Study 1 on Insomnia

Correlations among the variables are presented in Table 2. Repeated measures ANOVA with the imputed data revealed significant intervention by time interaction effect on insomnia severity ($F_{3,790.1259}$=7.434; P<.001), somatic presleep arousal ($F_{3,887.221}$=4.504; P=.004), cognitive presleep arousal ($F_{3,795.662}$=5.286; P=.001), dysfunctional beliefs and attitude about sleep ($F_{3,775.253}$=5.784; P<.001), mindfulness ($F_{3,749.436}$=3.590; P=.01), depression ($F_{3,838.002}$=3.938; P=.008), anxiety ($F_{3,845.675}$=4.554; P=.004), and mental well-being ($F_{3,826.162}$=3.482; P=.02), whereas the intervention by time interaction effect on functional adjustment was found to be nonsignificant ($F_{3,681.529}$=0.964; P=.41). Further analysis revealed that the intervention condition showed a better outcome on most of the outcome measures at postintervention, 1-month, and 3-month follow-up, compared with its respective wait-list control condition (see Table 3). Effect sizes are presented in Table 4.

XSL•FO
RenderX

Table 2. Correlations among variables of study 1 at baseline.

	ISI[a]	PSAS-S[b]	PSAS-C[c]	DBAS[d]	WSAS[e]	MAAS[f]	PHQ-9[g]	GAD-7[h]	WHO-5[i]
Insomnia Severity Index									
r	—[j]								
P value									
Presleep Arousal-Somatic									
r	0.420	—							
P value	<.001								
Presleep Arousal-Cognitive									
r	0.489	0.605	—						
P value	<.001	<.001							
16-item Dysfunctional Beliefs and Attitudes about Sleep									
r	0.401	0.330	0.387	—					
*P*value	<.001	<.001	<.001						
Work and Social Adjustment Scale									
r	0.365	0.330	0.311	0.475	—				
P value	<.001	<.001	<.001	<.001					
Mindful Attention Awareness Scale									
r	–0.237	–0.300	–0.405	–0.353	–0.335	—			
P value	<.001	<.001	<.001	<.001	<.001				
9-item Patient Health Questionnaire									
r	0.555	0.451	0.479	0.441	0.461	–0.562	—		
P value	<.001	<.001	<.001	<.001	<.001	<.001			
7-item Generalized Anxiety Disorder Scale									
r	0.461	0.483	0.590	0.449	0.436	–0.493	0.725	—	
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001		
5-item World Health Organization Well-being Index									
r	–0.407	–0.250	–0.295	–0.263	–0.230	0.245	–0.517	–0.407	—
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

[a]ISI: Insomnia Severity Index.

[b]PSAS-S: Presleep Arousal-Somatic.

[c]PSAS-C: Presleep Arousal-Cognitve.

[d]DBAS: Dysfunctional Beliefs and Attitudes about Sleep.

[e]WSAS: Work and Social Adjustment Scale.

[f]MAAS: Mindful Attention Awareness Scale.

[g]PHQ-9: 9-item Patient Health Questionnaire.

[h]GAD-7: 7-item Generalized Anxiety Disorder Scale.

[i]WHO-5: 5-item World Health Organization Well-being Index.

[j]Not applicable.

XSL•FO
RenderX

Table 3. Repeated measures analysis of variance for study 1: intent-to-treat analysis (N=333).

	Preinter-vention	Postinter-vention	1-month follow-up	3-month follow-up	Interaction effect		Condition effect		Time effect	
	Mean	Mean	Mean	Mean	*F (df)*	*P* value	*F (df)*	*P* value	*F (df)*	*P* value
Insomnia Severity Index					7.434 (3, 790.1259)	<.001	16.864 (1, 307.0944)	<.001	66.726 (3, 698.7932)	<.001
Intervention	15.32	10.38	10.75	10.35						
Wait-list	15.40	12.88	12.97	12.82						
Presleep Arousal-Somatic					4.504 (3, 887.221)	.004	6.544 (1, 312.570)	0.01	3.351 (3, 845.665)	.02
Intervention	6.77	5.07	5.84	5.59						
Wait-list	6.73	6.64	6.79	7.59						
Presleep Arousal-Cognitive					5.286 (3, 795.662)	.001	5.227 (1, 313.295)	.02	16.391 (3, 688.2804)	<.001
Intervention	15.08	11.76	11.95	11.07						
Wait-list	14.89	13.89	13.40	13.79						
Dysfunctional Beliefs and Attitudes about Sleep					5.784 (3, 775.253)	<.001	8.830 (1, 309.7971)	.003	9.690 (3, 646.1311)	<.001
Intervention	76.58	69.95	69.22	68.86						
Wait-list	76.81	75.40	75.44	75.75						
Mindful Attention Awareness Scale					3.590 (3, 749.436)	.01	1.072 (1, 318.1953)	.30	3.866 (3, 696.4202)	.009
Intervention	4.06	4.25	4.31	4.29						
Wait-list	4.15	4.11	4.18	4.13						
9-item Patient Health Questionnaire					3.938 (3, 838.0021)	.008	0.852 (1, 309.7324)	.36	11.400 (3, 785.1927)	<.001
Intervention	9.56	7.57	7.35	7.60						
Wait-list	8.84	8.49	7.97	8.44						
7-item Generalized Anxiety Disorder Scale					4.554 (3, 845.6747)	.004	2.132 (1, 316.0437)	.15	5.483 (3, 766.9396)	.001
Intervention	8.61	6.78	6.82	6.99						
Wait-list	8.06	8.10	7.73	7.98						
5-item World Health Organization Well-being Index					3.482 (3, 826.1617)	.02	3.221 (1, 304.5821)	.07	10.346 (3, 735.2119)	<.001
Intervention	31.40	38.42	38.14	42.17						
Wait-list	32.84	33.69	34.82	36.29						
Work and Social Adjustment Scale					0.964 (3, 681.5286)	.41	1.959 (1, 271.476)	.16	2.306 (3, 534.3949)	.08
Intervention	17.83	16.07	15.19	16.02						
Wait-list	17.68	17.17	16.81	17.63						

XSL•FO
RenderX

Table 4. Effect sizes of study 1.

	Postintervention versus preintervention				1-month follow-up versus preintervention				3-month follow-up versus preintervention			
	t (df)	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI
Insomnia Severity Index												
Intervention	13.143 (5443.323)	<.001	1.11	0.92 to 1.31	10.392 (2254.874)	<.001	0.97	0.78 to 1.16	11.762 (1547.873)	<.001	1.04	0.85 to 1.23
Wait-list	7.038 (1622.922)	<.001	1.06	0.78 to 1.34	5.606 (894.225)	<.001	0.50	0.34 to 0.65	6.416 (1233.278)	<.001	0.53	0.38 to 0.67
Presleep Arousal-Somatic												
Intervention	4.451 (15081.001)	<.001	0.34	0.19 to 0.49	2.184 (6574.37)	.03	0.19	0.03 to 0.34	2.637 (4766.154)	.008	0.23	0.07 to 0.40
Wait-list	0.267 (2637.896)	.79	0.04	–0.21 to 0.28	–0.143 (3473.035)	.89	0.01	–0.16 to 0.14	–1.977 (2875.841)	.048	–0.16	–0.03 to –0.01
Presleep Arousal-Cognitive												
Intervention	6.794 (8818.592)	<.001	0.49	0.35 to 0.64	5.234 (1371.045)	<.001	0.45	0.30 to 0.61	6.884 (1663.291)	<.001	0.58	0.43 to 0.74
Wait-list	2.086 (1186.225)	.04	0.28	0.05 to 0.50	2.797 (1085.731)	.005	0.21	0.08 to 0.33	2.092 (2344.107)	.04	0.15	0.02 to 0.27
16-item Dysfunctional Beliefs and Attitudes about Sleep												
Intervention	6.397 (3434.157)	<.001	0.43	0.30 to 0.56	5.976 (1548.938)	<.001	0.45	0.31 to 0.59	5.662 (1459.862)	<.001	0.45	0.31 to 0.59
Wait-list	1.349 (1526.281)	.18	0.18	0.05 to 0.42	1.143 (564.757)	.25	0.09	–0.03 to 0.21	0.863 (873.599)	.39	–0.07	–0.06 to 0.19
Mindful Attention Awareness Scale												
Intervention	–3.659 (3932.992)	<.001	0.23	0.12 to 0.35	–4.724 (1464.818)	<.001	0.31	0.20 to 0.43	–3.830 (1361.077)	<.001	0.28	0.15 to 0.40
Wait-list	0.668 (1370.984)	.50	–0.07	–0.27 to 0.13	–0.670 (975.659)	.50	–0.05	–0.15 to 0.05	0.207 (1054.694)	.84	–0.01	–0.13 to 0.10
9-item Patient Health Questionnaire												
Intervention	5.456 (5123.518)	<.001	0.43	0.28 to 0.58	5.864 (2734.949)	<.001	0.48	0.33 to 0.64	5.003 (2211.678)	<.001	0.41	0.26 to 0.56
Wait-list	1.014 (2969.454)	.31	0.14	–0.11 to 0.39	2.361 (1450.462)	.02	0.18	0.05 to 0.31	1.049 (2210.411)	.29	0.08	–0.05 to 0.22
7-item Generalized Anxiety Disorder Scale												
Intervention	5.022 (8639.716)	<.001	0.38	0.23 to 0.53	4.685 (2436.844)	<.001	0.37	0.23 to 0.52	4.308 (2432.34)	<.001	0.33	0.19 to 0.47
Wait-list	–0.102 (2277.929)	.92	–0.02	–0.26 to 0.23	0.832 (1421.884)	.41	0.07	–0.07 to 0.20	0.225 (2365.867)	.82	0.02	–0.11 to 0.15
5-item World Health Organization Well-being Index												
Intervention	–4.387 (2744.74)	<.001	0.36	0.21 to 0.51	–4.417 (2159.869)	<.001	0.35	0.21 to 0.49	–6.318 (1335.566)	<.001	0.55	0.39 to 0.71
Wait-list	–0.523 (2592.251)	.60	–0.09	–0.40 to 0.22	–1.303 (1338.226)	.19	–0.11	–0.24 to 0.03	–2.088 (1592.958)	.04	–0.19	–0.34 to –0.03
Work and Social Adjustment Scale												
Intervention	2.107 (822.965)	.04	0.21	0.05 to 0.37	3.233 (743.299)	.001	0.31	0.16 to 0.47	2.257 (778.492)	.02	0.21	0.06 to 0.37
Wait-list	0.601 (674.646)	.55	0.12	–0.20 to 0.45	0.970 (394.396)	.33	0.1	–0.05 to 0.26	0.063 (688.404)	.95	0	–0.15 to 0.16

XSL•FO
RenderX

Mediation Model

One-Month Follow-up Results

The results showed that all 4 models had satisfactory model fit. In addition, at 1-month follow-up assessment, the indirect effects of condition on all primary and secondary outcomes through dysfunctional beliefs and attitudes about sleep were significant, except mental well-being. The indirect effects of condition through mindfulness were significant on cognitive presleep arousal.

Three-Month Follow-up Results

At 3-month follow-up assessment, all 4 models showed satisfactory model fit. The indirect effects of condition on the primary and secondary outcomes through dysfunctional beliefs and attitudes about sleep were all significant. However, the indirect effects of condition through mindfulness were only significant on cognitive and somatic presleep arousal. Table 5 shows a summary of the model fit, standardized path coefficients, standard errors, indirect effects, total effects, and model fits of the 4 models. Figure 4 shows the mediation models of study 1.

XSL•FO
RenderX

Table 5. Standardized path coefficients, standard errors, indirect effects, total effects, and model fits of the mediation analyses in study 1.

Model, dependent variables	a (Condition → Mindfulness)	b (Condition → Dysfunctional Beliefs and Attitude about Sleep Scale)	c (Mindfulness → dependent variable[c])	d (Dysfunctional Beliefs and Attitude about Sleep Scale → dependent variable)	e (Condition → dependent variable)	Indirect effect through mindfulness (a*c)	Indirect effect through Dysfunctional Beliefs and Attitude about Sleep Scale (b*d)	Total effect
1a, Primary outcomes at 1-month follow-up[a]								
Insomnia Severity Index	0.155 (0.042), *P*<.001	–0.224 (0.041), *P*<.001	–0.109 (0.053), *P*=.04	0.212 (0.058), *P*<.001	–0.262 (0.051), *P*<.001	–0.017 (0.010), *P*=.08	–0.047 (0.015), *P*=.002	–0.327 (0.049), *P*<.001
Presleep Arousal-Somatic	N/A[b]	N/A	–0.119 (0.052), *P*=.02	0.164 (0.055), *P*=.003	–0.124 (0.051), *P*=.02	–0.018 (0.010), *P*=.05	–0.037 (0.014), *P*=.008	–0.179 (0.050), *P*<.001
Presleep Arousal-Cognitive	N/A	N/A	–0.169 (0.048), *P*<.001	0.194 (0.051), *P*<.001	–0.116 (0.047), *P*=.01	–0.026 (0.010), *P*=.01	–0.043 (0.014), *P*=.001	–0.186 (0.047), *P*<.001
1b, Primary outcomes at 3-month follow-up[c]								
Insomnia Severity Index	0.155 (0.042), *P*<.001	–0.224 (0.041), *P*<.001	–0.050 (0.051), *P*=.33	0.249 (0.053), *P*<.001	–0.281 (0.048), *P*<.001	–0.008 (0.008), *P*=.35	–0.056 (0.016), *P*<.001	–0.344 (0.047), *P*<.001
Presleep Arousal-Somatic	N/A	N/A	–0.175 (0.050), *P*<.001	0.180 (0.054) *P*=.001	–0.201 (0.048), *P*<.001	–0.027 (0.011), *P*=.01	–0.040 (0.014), *P*=.004	–0.268 (0.046), *P*<.001
Presleep Arousal-Cognitive	N/A	N/A	–0.164 (0.047), *P*<.001	0.210 (0.049), *P*<.001	–0.217 (0.044), *P*<.001	–0.025 (0.010), *P*=.01	–0.047 (0.014), *P*=.001	–0.289 (0.043), *P*<.001
2a, Secondary outcomes at 1-month follow-up[d]								
Work and Social Adjustment Scale	0.119 (0.050), *P*=.02	–0.216 (0.048), *P*<.001	–0.120 (0.066), *P*=.07	0.272 (0.066), *P*<.001	–0.203 (0.060), *P*=.001	–0.014 (0.010), *P*=.16	–0.059 (0.019), *P*=.002	–0.277 (0.060), *P*<.001
9-item Patient Health Questionnaire	N/A	N/A	–0.0140 (0.058), *P*=.02	0.370 (0.058), *P*<.001	–0.059 (0.059), *P*=.32	–0.017 (0.010), *P*=.09	–0.080 (0.022), *P*<.001	–0.156 (0.061), *P*=.01
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.160 (0.064), *P*=.01	0.229 (0.065), *P*<.001	–0.106 (0.061), *P*=.08	–0.019 (0.011), *P*=.09	–0.049 (0.018), *P*=.005	–0.174 (0.060), *P*=.004
5-item World Health Organization Well-being Index	N/A	N/A	0.133 (0.063), *P*=.03	–0.113 (0.068), *P*=.09	0.154 (0.060), *P*=.01	0.016 (0.010), *P*=.12	0.024 (0.016), *P*=.12	0.194 (0.059), *P*=.001
2b, Secondary outcomes at 3-month follow-up[e]								
Work and Social Adjustment Scale	0.119 (0.050), *P*=.02	–0.216 (0.048), *P*<.001	–0.191 (0.060), *P*=.001	0.346 (0.064), *P*<.001	–0.157 (0.059), *P*=.008	–0.023 (0.012), *P*=.06	–0.075 (0.021), *P*<.001	–0.254 (0.060), *P*<.001
9-item Patient Health Questionnaire	N/A	N/A	–0.196 (0.061), *P*=.001	0.327 (0.061), *P*<.001	–0.088 (0.059), *P*=.13	–0.023 (0.012), *P*=.06	–0.071 (0.021), *P*=.001	–0.182 (0.059), *P*=.002
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.148 (0.064), *P*=.02	0.201 (0.064), *P*=.002	–0.110 (0.059), *P*=.06	–0.018 (0.011), *P*=.11	–0.043 (0.017), *P*=.009	–0.171 (0.059), *P*=.004
5-item World Health Organization Well-being Index	N/A	N/A	0.071 (0.071), *P*=.32	–0.158 (0.071), *P*=.03	0.084 (0.063), *P*=.18	0.008 (0.009), *P*=.37	0.034 (0.017), *P*=.048	0.127 (0.061), *P*=.04

[a]Model fit: χ^2_{20}=31.3; *P*=.051; comparative fit index=0.988; Tucker–Lewis index=0.976; standardized root mean squared residual=0.029; root mean

XSL•FO
RenderX

square error of approximation=0.041

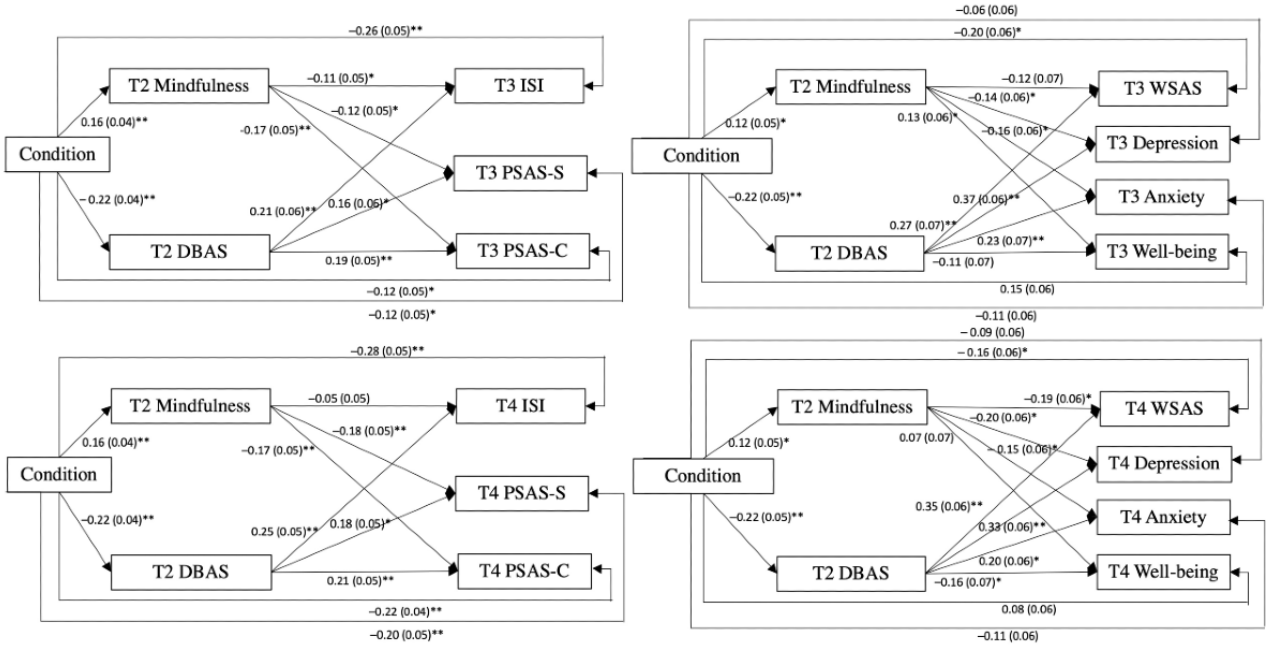
[b]N/A: not applicable.

[c]Model fit: χ^2_{20}=39.8; *P*=.005; comparative fit index=0.981; Tucker–Lewis index=0.961; standardized root mean squared residual=0.033; root mean square error of approximation=0.055.

[d]Model fit: χ^2_{30}=53.2; *P*=.006; comparative fit index=0.970; Tucker–Lewis index=0.942; standardized root mean squared residual=0.044; root mean square error of approximation=0.057.

[e]Model fit: χ^2_{30}=40.5; *P*=.09; comparative fit index=0.985; Tucker–Lewis index=0.971; standardized root mean squared residual=0.032; root mean square error of approximation=0.038.

Figure 4. Path models of study 1. Baseline variables were controlled. For the purpose of clarity, the controlled variables and the covariance of the residuals between variables are not depicted. DBAS: Dysfunctional Beliefs and Attitudes about Sleep; ISI: Insomnia Severity Index; PSAS-C: Presleep Arousal-Cognitive; PSAS-S: Presleep Arousal-Somatic; T2: postassessment; T3: 1-month follow-up assessment; T4: 3-month follow-up assessment; WSAS: Work and Social Adjustment Scale. *P<.05, **P<.001.



Study 2 on Pain

Correlations among variables are presented in Table 6. Repeated measures ANOVA results showed significant intervention condition by time interaction effect on pain acceptance ($F_{3,535.514}$=4.186; *P*=.006), and pain catastrophizing ($F_{3,558.550}$=3.179; *P*=.02), mindfulness ($F_{3,597.884}$=3.003; *P*=.03), depression ($F_{3,592.642}$=2.781; *P*=.04), and mental well-being ($F_{3,593.052}$=3.762; *P*=.01). Meanwhile, interaction effects were found to be nonsignificant on subjective intensity of pain ($F_{3,591.397}$=0.464; *P*=.71), functional adjustment ($F_{3,541.571}$=1.720; *P*=.16), and anxiety ($F_{3,569.171}$=2.230; *P*=.08). Further analysis showed that the significant interaction effects were largely maintained across postintervention, 1-month, and 3-month follow-ups. Participants in the intervention condition reported better outcomes on most of the measures (See Table 7). Effect sizes are presented in Table 8.

XSL•FO
RenderX

Table 6. Correlations among variables of study 2 at baseline.

	VAS[a]	CPAQ[b]	PCS[c]	WSAS[d]	MAAS[e]	PHQ-9[f]	GAD-7[g]	WHO-5[h]
Visual Analogue Scale								
r	—[i]							
P value								
Chronic Pain Acceptance Questionnaire								
r	–0.405	—						
P value	<.001							
Pain Catastrophizing Scale								
r	0.449	–0.642	—					
P value	<.001	<.001						
Work and Social Adjustment Scale								
r	0.438	–0.603	0.635	—				
P value	<.001	<.001	<.001					
Mindful Attention Awareness Scale								
r	–0.117	0.276	–0.418	–0.344	—			
P value	0.07	<.001	<.001	<.001				
9-item Patient Health Questionnaire								
r	0.214	–0.325	0.483	0.371	–0.627	—		
P value	<.001	<.001	<.001	<.001	<.001			
7-item Generalized Anxiety Disorder Scale								
r	0.174	–0.327	0.470	0.434	–0.554	0.751	—	
P value	0.007	<.001	<.001	<.001	<.001	<.001		
5-item World Health Organization Well-being Index								
r	–0.218	0.391	–0.466	–0.342	0.426	–0.613	–0.598	—
P value	<.001	<.001	<.001	<.001	<.001	<.001	<.001	

[a]VAS: Visual Analogue Scale.

[b]CPAQ: Chronic Pain Acceptance Questionnaire.

[c]PCS: Pain Catastrophizing Scale.

[d]WSAS: Work and Social Adjustment Scale.

[e]MAAS: Mindful Attention Awareness Scale.

[f]PHQ-9: 9-item Patient Health Questionnaire.

[g]GAD-7: 7-item Generalized Anxiety Disorder Scale.

[h]WHO-5: 5-item World Health Organization Well-being Index.

[i]Not applicable.

XSL•FO
RenderX

Table 7. Repeated measures analysis of variance for study 2: intent-to-treat analysis.

	Preinterven-tionn	Postinterven-tion	1-month follow-up	3-month follow-up	Interaction effect		Condition effect		Time effect	
	Mean	Mean	Mean	Mean	*F (df)*	*P* value	*F (df)*	*P* value	*F (df)*	*P* value
Visual Analogue Scale					0.464 (3, 591.3966)	.71	6.525 (1, 210.2334)	.01	9.533 (3, 552.7345)	<.001
Intervention	49.36	40.09	39.28	40.91						
Wait-list	52.74	45.56	46.09	47.92						
Chronic Pain Acceptance Questionnaire					4.186 (3, 535.5144)	.006	1.015 (1, 216.9613)	.32	6.591 (3, 508.6055)	<.001
Intervention	58.96	64.41	64.49	63.85						
Wait-list	60.88	63.16	60.57	61.18						
Pain Catastrophizing Scale					3.179 (3, 558.5502)	.02	3.370 (1, 215.56)	.07	5.365 (3, 535.191)	.001
Intervention	2.84	2.52	2.56	2.53						
Wait-list	2.81	2.75	2.75	2.79						
Mindful Attention Awareness Scale					3.003 (3, 597.8836)	.03	5.068 (1, 220.251)	.03	1.958 (3, 511.7978)	.12
Intervention	4.19	4.37	4.39	4.47						
Wait-list	4.15	4.12	4.19	4.11						
9-item Patient Health Questionnaire					2.781 (3, 592.6418)	.04	0.062	.80	1.445 (3, 533.4394)	.23
Intervention	7.60	6.50	6.95	10.88						
Wait-list	7.77	8.25	8.39	12.75						
7-item Generalized Anxiety Disorder Scale					2.230 (3, 569.1707)	.08	6.244 (1, 218.5466)	.01	0.777 (3, 559.0639)	.51
Intervention	6.80	5.82	6.33	6.30						
Wait-list	7.17	7.66	8.08	7.96						
5-item World Health Organization Well-being Index					3.762 (3, 593.0518)	.01	1.886 (1, 216.6309)	.17	2.281 (3, 578.2624)	.08
Intervention	39.93	48.11	46.50	45.79						
Wait-list	41.85	39.55	41.83	43.92						
Work and Social Adjustment Scale					1.720 3, 541.5711)	.16	0.535 (1, 200.4921)	.47	4.790 (3, 437.0823)	.003
Intervention	17.79	13.42	15.32	13.88						
Wait-list	16.40	14.84	16.64	15.43						

XSL•FO
RenderX

Table 8. Effect sizes of study 2.

	Postintervention versus preintervention				1-month follow-up versus preintervention				3-month follow-up versus preintervention			
	t (df)	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI
Visual Analogue Scale												
Intervention	4.236 (3020.777)	<.001	0.47	0.26 to 0.67	4.042 (2266.412)	<.001	0.48	0.27 to 0.71	3.230 (1716.028)	.001	0.40	0.18 to 0.62
Wait-list	3.351 (1895.399)	.001	0.35	0.16 to 0.54	2.929 (1427.677)	.003	0.32	0.13 to 0.50	2.04 (1866.139)	.04	0.23	0.03 to 0.42
Chronic Pain Acceptance Questionnaire												
Intervention	–5.597 (3163.368)	<.001	0.45	0.30 to 0.61	–4.763 (1184.051)	<.001	0.45	0.28 to 0.62	–3.467 (822.627)	.001	0.38	0.57 to 0.57
Wait-list	–2.187 (1276.259)	.03	0.18	0.04 to 0.32	0.297 (1075.033)	.77	–0.03	–0.16 to 0.11	–0.260 (992.647)	.79	0.02	–0.12 to 0.17
Pain Catastrophizing Scale												
Intervention	4.801 (2844.541)	<.001	0.42	0.26 to 0.59	3.949 (1366.464)	<.001	0.38	0.21 to 0.55	3.616 (1387.779)	<.001	0.38	0.20 to 0.57
Wait-list	0.920 (1543.763)	.36	0.08	–0.07 to 0.22	0.872 (1408.471)	.38	0.08	–0.06 to 0.22	0.286 (1009.163)	.78	0.03	–0.12 to 0.18
Mindful Attention Awareness Scale												
Intervention	–2.593 (4982.628)	.01	0.24	0.07 to 0.41	–2.726 (2285.433)	.006	0.27	0.09 to 0.45	–3.255 (1058.504)	.001	0.35	0.17 to 0.54
Wait-list	0.391 (1549.393)	.69	–0.04	–0.17 to 0.10	–0.579 (1657.309)	.56	0.05	–0.09 to 0.18	0.490 (1088.182)	.62	–0.04	–0.18 to 0.09
9-item Patient Health Questionnaire												
Intervention	2.832 (3259.39)	.005	0.25	0.09 to 0.42	1.504 (1163.214)	.13	0.14	0.02 to 0.29	0.880 (799.257)	.38	0.09	–0.07 to 0.25
Wait-list	–1.261 (2701.169)	.21	–0.09	–0.21 to 0.04	–1.360 (1804.568)	.17	–0.12	–0.26 to 0.03	–2.615 (1727.604)	.009	–0.23	–0.38 to –0.07
7-item Generalized Anxiety Disorder Scale												
Intervention	2.537 (1765.07)	.01	0.24	0.08 to 0.41	1.088 (717.43)	.28	0.11	0.05 to 0.27	1.118 (1134.925)	.26	0.11	0.06 to 0.29
Wait-list	–1.267 (2817.28)	.21	–0.09	–0.22 to 0.04	–2.057 (1693.877)	.04	–0.17	–0.31 to –0.03	–1.634 (2480.039)	.10	–0.14	–0.29 to 0.01
5-item World Health Organization Well-being Index												
Intervention	–4.207 (4037.397)	<.001	0.38	0.21 to 0.56	–3.530 (1700.367)	<.001	0.31	0.15 to 0.46	–2.231 (1714.783)	.03	0.27	0.06 to 0.48
Wait-list	1.169 (1678.009)	.24	–0.11	–0.27 to 0.05	0.007 (1799.488)	.99	0	–0.16 to 0.16	–1.028 (1669.093)	.30	0.09	0.06 to 0.25
Work and Social Adjustment Scale												
Intervention	4.328 (690.693)	<.001	0.49	0.30 to 0.68	2.241 (614.852)	.03	0.27	0.08 to 0.45	3.377 (712.65)	.001	0.43	0.22 to 0.64
Wait-list	1.569 (671.049)	.12	0.17	0.00 to 0.34	–0.213 (834.573)	.83	–0.03	–0.22 to 0.16	0.935 (688.875)	.35	0.11	–0.07 to 0.29

Mediation Model

One-Month Follow-up Assessment

Results showed that condition had significant indirect effects on pain intensity and functional adjustment through pain catastrophizing, whereas condition showed significant indirect effects on depression and anxiety through mindfulness. All models showed satisfactory model fit.

Three-Month Follow-up Assessment

Similar to results at 1-month follow-up, the indirect effects of condition through pain catastrophizing on pain intensity and functional adjustment at 3-month follow-up were significant,

XSL•FO
RenderX

and the indirect effects of condition through mindfulness on depression and anxiety were also significant. A summary of the model fit and standardized coefficients are shown in Table 9 and Figure 5.

Table 9. Standardized path coefficients, standard errors, indirect effects, total effects, and model fit of the mediation analyses in study 2.

Dependent variables	a (Condition → Mindfulness)	b (Condition → Pain Catastrophizing Scale)	c (Mindfulness → dependent variable)	d (Pain Catastrophizing Scale → dependent variable)	e (Condition → dependent variable)	Indirect effect through mindfulness (a*c)	Indirect effect through Pain Catastrophizing Scale (b*d)	Total effect
Primary outcomes at 1-month follow-up[a]								
Pain intensity	0.141 (0.056), *P*=.01	–0.188 (0.057), *P*=.001	–0.061 (0.071), *P*=.39	0.350 (0.072), *P*<.001	–0.096 (0.063), *P*=.13	–0.009 (0.011), *P*=.42	–0.066 (0.025), *P*=.007	–0.171 (0.064), *P*=.007
Primary outcomes at 3-month follow-up[b]								
Pain intensity	0.141 (0.056), *P*=.01	–0.188 (0.057), *P*=.001	–0.099 (0.076), *P*=.19	0.281 (0.077), *P*<.001	–0.164 (0.065), *P*=.01	–0.014 (0.012), *P*=.26	–0.053 (0.022), *P*=.02	–0.231 (0.064), *P*<.001
Secondary outcomes at 1-month follow-up[c]								
Work and Social Adjustment Scale	0.178 (0.065), *P*=.006	–0.181 (0.065), *P*=.005	–0.042 (0.078), *P*=.59	0.408 (0.083), *P*<.001	–0.107 (0.074), *P*=.15	–0.008 (0.015), *P*=.61	–0.074 (0.032), *P*=.02	–0.189 (0.074), *P*=.01
9-item Patient Health Questionnaire	N/A[d]	N/A	–0.244 (0.074), *P*=.001	0.103 (0.075), *P*=.17	–0.121 (0.065), *P*=.06	–0.044 (0.021), *P*=.04	–0.019 (0.016), *P*=.23	–0.184 (0.065), *P*=.005
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.226 (0.071), *P*=.002	0.099 (0.070), *P*=.16	–0.161 (0.064), *P*=.012	–0.041 (0.020), *P*=.045	–0.018 (0.015), *P*=.22	–0.219 (0.063), *P*=.001
5-item World Health Organization Well-being Index	N/A	N/A	0.228 (0.084), *P*=.007	0.068 (0.084), *P*=.42	0.185 (0.070), *P*=.008	0.041 (0.022), *P*=.06	–0.012 (0.016), *P*=.45	0.214 (0.068), *P*=.002
Secondary outcomes at 3-month follow-up[e]								
Work and Social Adjustment Scale	0.178 (0.065) *P*=.006	–0.181 (0.065), *P*=.005	–0.042 (0.080), *P*=.59	0.472 (0.082), *P*<.001	–0.169 (0.074), *P*=.02	–0.008 (0.015), *P*=.61	–0.086 (0.035), *P*=.02	–0.263 (0.073), *P*<.001
9-item Patient Health Questionnaire	N/A	N/A	–0.265 (0.073), *P*<.001	0.149 (0.071), *P*=.04	–0.232 (0.065), *P*<.001	–0.047 (0.022), *P*=.03	–0.027 (0.017), *P*=.10	–0.306 (0.064), *P*<.001
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.232 (0.074), *P*=.002	0.192 (0.071), *P*=.007	–0.192 (0.066), *P*=.004	–0.042 (0.021), *P*=.04	–0.035 (0.019), *P*=.06	–0.268 (0.066), *P*<.001
5-item World Health Organization Well-being Index	N/A	N/A	0.152 (0.085), *P*=.07	0.007 (0.084), *P*=.93	0.140 (0.076), *P*=.07	0.027 (0.018), *P*=.14	–0.001 (0.016), *P*=.93	0.165 (0.074), *P*=.03

[a]Model fit: χ^2_6=5.5; *P*=.49; comparative fit index=1.00; Tucker–Lewis index=1.005; standardized root mean squared residual=0.034; root mean square error of approximation<.001

[b]Model fit: χ^2_6=6.2; *P*=.40; comparative fit index=0.999; Tucker–Lewis index=0.998; standardized root mean squared residual=0.34; root mean square error of approximation=0.011.

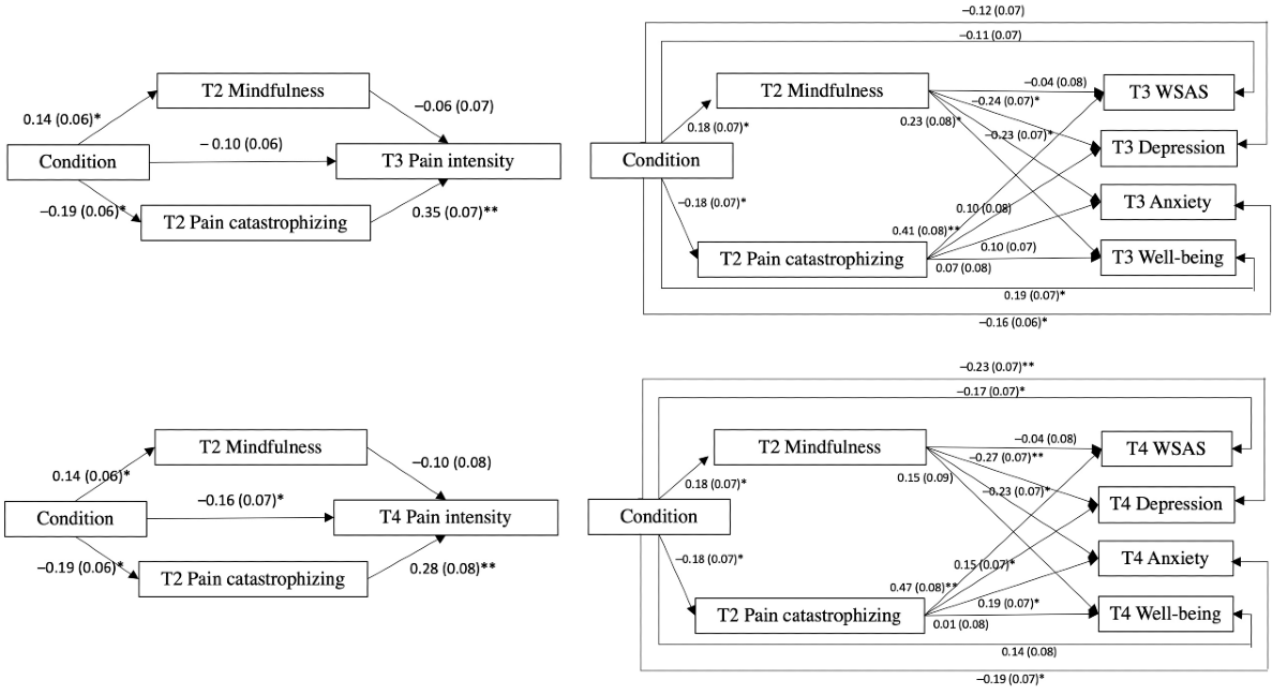
[c]Model fit: χ^2_{30}=29.2; *P*=.51; comparative fit index=1.00; Tucker–Lewis index=1.003; standardized root mean squared residual=0.04; root mean square error of approximation<.001.

[d]N/A: not applicable.

[e]Model fit: χ^2_{30}=25.9; *P*=.68; comparative fit index=1.00; Tucker–Lewis index=1.013; standardized root mean squared residual=0.040; root mean square error of approximation<.001.

XSL•FO
RenderX

Figure 5. Path models of study 2. Baseline variables were controlled. For the purpose of clarity, the controlled variables and the covariance of the residuals between variables are not depicted. T2: postassessment; T3: 1-month follow-up assessment; T4: 3-month follow-up assessment; WSAS: Work and Social Adjustment Scale. *P<.05, **P<.001.



Study 3 on Dysregulated Eating

Correlations among variables are presented in Table 10. Results of repeated measures ANOVA showed significant intervention condition by time interaction effect on uncontrolled eating ($F_{3,912.364}$=3.041; P=.03), emotional eating ($F_{3,944.138}$=4.294; P=.005), mindfulness ($F_{3,854.310}$=2.670; P=.05), and mental well-being ($F_{3,882.777}$=4.457; P=.004). Meanwhile, analyses on other measures showed nonsignificant results. For the outcome with significant results, further analyses showed that significant group difference was maintained and the intervention condition reported better outcomes at postintervention, 1-month, and 3-month follow-up (See Table 11). Effect sizes are presented in Table 12.

XSL•FO
RenderX

Table 10. Correlations among variables of study 3 at baseline.

	TFEQ-UE[a]	TFEQ-CR[b]	TFEQ-EE[c]	PFS[d]	MAAS[e]	PHQ-9[f]	GAD-7[g]	WHO-5[h]
Three-Factor Eating Questionnaire-Uncontrolled Eating								
r	—[i]							
P value								
Three-Factor Eating Questionnaire-Cognitive Restraint								
r	–0.038	—						
P value	.48							
Three-Factor Eating Questionnaire-Emotional Eating								
r	0.643	–0.009	—					
P value	<.001	.87						
Power of Food Scale								
r	0.708	–0.048	0.538	—				
P value	<.001	.37	<.001					
Mindful Attention Awareness Scale								
r	–0.368	0.110	–0.423	–0.358	—			
P value	<.001	.04	<.001	<.001				
9-item Patient Health Questionnaire								
r	0.251	–0.116	0.373	0.288	–0.626	—		
P value	<.001	.03	<.001	<.001	<.001			
7-item Generalized Anxiety Disorder Scale								
r	0.200	–0.027	0.304	0.206	–0.564	0.796	—	
P value	<.001	.62	<.001	<.001	<.001	<.001		
5-item World Health Organization Well-being Index								
r	–0.086	0.107	–0.243	–0.102	0.46	–0.592	–0.578	—
P value	0.11	.05	<.001	.06	<.001	<.001	<.001	

[a]TFEQ-UE: Three-Factor Eating Questionnaire-Uncontrolled Eating.

[b]TFEQ-CR: Three-Factor Eating Questionnaire-Cognitive Restraint.

[c]TFEQ-EE: Three-Factor Eating Questionnaire-Emotional Eating.

[d]PFS: Power of Food Scale.

[e]MAAS: Mindful Attention Awareness Scale.

[f]PHQ-9: 9-item Patient Health Questionnaire.

[g]GAD-7: 7-item Generalized Anxiety Disorder Scale.

[h]WHO-5: 5-item World Health Organization Well-being Index.

[i]Not applicable.

XSL•FO
RenderX

Table 11. Repeated measures analysis of variance for study 3: intent-to-treat analysis (N=351).

	Preintervention	Postintervention	1-month follow-up	3-month follow-up	Interaction effect		Condition effect		Time effect	
	Mean	Mean	Mean	Mean	*F (df)*	*P* value	*F (df)*	*P* value	*F (df)*	*P* value
Three-Factor Eating Questionnaire-Uncontrolled Eating					3.041 (3, 912.364)	.03	10.275 (1, 325.2121)	.002	21.132 (3, 841.623)	<.001
Intervention	20.95	19.21	18.97	18.70						
Wait-list	21.53	21.06	20.40	20.37						
Three-Factor Eating Questionnaire-Cognitive Restraint					0.135 (3, 875.081)	.94	0.086 (1, 303.5693)	.77	3.299 (3, 793.3058)	.02
Intervention	9.21	9.51	9.59	9.50						
Wait-list	9.12	9.52	9.52	9.35						
Three-Factor Eating Questionnaire-Emotional Eating					4.294 (3, 944.138)	.005	7.781 (1, 324.9344)	.006	4.792 (3, 866.233)	.003
Intervention	7.58	6.86	6.88	6.84						
Wait-list	7.61	7.58	7.52	7.57						
Power of Food Scale					0.771 (3, 1031.464)	.51	19.932 (1, 343.1272)	<.001	23.943 (3, 1030.6991)	<.001
Intervention	43.64	40.38	40.32	39.16						
Wait-list	47.22	45.89	44.56	43.98						
Mindful Attention Awareness Scale					2.670 (3, 854.310)	.05	5.346 (1, 310.4349)	.02	7.599 (3, 793.6769)	<.001
Intervention	4.05	4.29	4.31	4.33						
Wait-list	4.00	4.05	4.08	4.16						
9-item Patient Health Questionnaire					1.241 (3, 914.466)	.29	2.715 (1, 320.5567)	.10	4.236 (3, 826.3964)	.006
Intervention	7.61	6.72	6.27	6.37						
Wait-list	7.74	7.75	6.99	7.19						
7-item Generalized Anxiety Disorder Scale					1.75 (3, 900.082)	.16	1.557 (1, 323.1843)	.21	2.048 (3, 825.7671)	.11
Intervention	6.99	5.87	5.93	5.89						
Wait-list	6.86	6.89	6.50	6.57						
5-item World Health Organization Well-being Index					4.457 (3, 882.777)	.004	1.890 (1, 320.3858)	.17	2.178 (3, 789.4446)	.09
Intervention	42.97	50.12	48.36	48.68						
Wait-list	44.94	43.92	46.67	44.78						

XSL•FO
RenderX

Table 12. Effect sizes of study 3.

	Postintervention versus Preintervention				1-month follow-up versus Preintervention				3-month follow-up versus Preintervention			
	t (df)	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI	*t (df)*	*P* value	Cohen *d*	CI
Three-Factor Eating Questionnaire-Uncontrolled Eating												
Intervention	5.186 (3817.101)	<.001	0.36	0.23 to 0.50	6.105 (3639.106)	<.001	0.45	0.31 to 0.59	5.861 (2319.303)	<.001	0.46	0.32 to 0.61
Wait-list	1.930 (4614.144)	.05	0.12	0.01 to 0.23	2.893 (3664.02)	.004	0.20	0.07 to 0.33	3.170 (2504.465)	.002	0.24	0.12 to 0.37
Three-Factor Eating Questionnaire-Cognitive Restraint												
Intervention	–1.643 (2879.543)	.005	0.13	–0.01 to 0.28	–1.801 (1687.066)	.07	0.16	0.01 to 0.31	–1.366 (1069.425)	.17	0.12	–0.03 to 0.27
Wait-list	–2.440 (3497.992)	.02	0.18	0.05 to 0.31	–2.106 (2073.137)	.04	0.17	0.03 to 0.30	–1.303 (2226.666)	.19	0.10	–0.03 to 0.23
Three-Factor Eating Questionnaire-Emotional Eating												
Intervention	4.721 (1990.064)	.001	0.33	0.20 to 0.45	4.393 (2491.901)	<.001	0.32	0.19 to 0.46	4.431 (3790.093)	<.001	0.35	0.20 to 0.49
Wait-list	0.197 (3848.245)	.84	0.01	–0.11 to 0.14	0.111 (2995.514)	.91	0.00	–0.13 to 0.14	0.320 (2534.89)	.75	0.02	–0.11 to 0.15
Power of Food Scale												
Intervention	4.381 (50588.352)	<.001	0.30	0.16 to 0.43	4.367 (121149.766)	<.001	0.34	0.18 to 0.49	4.769 (113062.726)	<.001	0.46	0.32 to 0.61
Wait-list	2.739 (112506.843)	.006	0.17	0.05 to 0.30	3.570 (27839.166)	<.001	0.24	0.11 to 0.36	5.298 (62137.205)	<.001	0.35	0.22 to 0.48
Mindful Attention Awareness Scale												
Intervention	–3.686 (1565.495)	.001	0.27	0.14 to 0.39	–4.124 (1323.734)	<.001	0.29	0.17 to 0.41	–4.180 (1425.676)	<.001	0.12	–0.03 to 0.27
Wait-list	–0.940 (5467.915)	.35	0.07	–0.07 to 0.20	–0.524 (2393.459)	.60	0.03	–0.09 to 0.16	–1.551 (2324.842)	.12	0.12	–0.02 to 0.26
9-item Patient Health Questionnaire												
Intervention	2.321 (3106.031)	.12	0.17	0.04 to 0.30	2.948 (3308.172)	.003	0.25	0.10 to 0.40	2.868 (2623.881)	.004	0.35	0.20 to 0.49
Wait-list	–0.099 (6576.707)	.92	0.01	–0.15 to 0.14	1.502 (2597.217)	.13	0.12	–0.02 to 0.26	0.815 (2368.508)	.42	0.07	0.08 to 0.21
7-item Generalized Anxiety Disorder Scale												
Intervention	3.096 (2600.162)	.047	0.21	0.09 to 0.34	2.373 (2117.166)	.02	0.18	0.05 to 0.32	2.646 (2359.088)	.008	0.39	0.23 to 0.55
Wait-list	–0.171 (6545.757)	.86	–0.01	–0.15 to 0.13	0.344 (4111.079)	.73	0.03	–0.12 to 0.17	0.347 (2103.419)	.73	0.03	–0.12 to 0.18
5-item World Health Organization Well-being Index												
Intervention	–4.243 (3063.119)	.02	0.33	0.19 to 0.48	–2.920 (1302.715)	.004	0.22	0.09 to 0.36	–2.651 (1620.939)	.008	0.33	0.19 to 0.46
Wait-list	0.885 (4256.998)	.38	–0.07	–0.21 to 0.07	–0.487 (1512.967)	.63	0.04	–0.10 to 0.17	0.628 (2220.244)	.53	0.06	–0.21 to 0.10

Mediation Model

One-Month Follow-up Results

Results showed that at 1-month follow-up assessments, condition showed significant indirect effects on uncontrolled eating and emotional eating through power of food. Condition also showed indirect effects on depression and anxiety through mindfulness.

Three-Month Follow-up Results

At 3-month follow-up assessments, condition showed significant indirect effects on uncontrolled eating and emotional eating through power of food, whereas it showed significant indirect

XSL•FO
RenderX

effect on depression through mindfulness. A summary of the model fit, standardized coefficients, indirect effects, and total effects is shown in Table 13 and Figure 6.

XSL•FO
RenderX

Table 13. Standardized path coefficients, indirect effects, total effects, and model fit of the mediation analyses in study 3.

Dependent variables	a (Condition → Mindfulness)	b (Condition → Power of Food Scale)	c (Mindfulness → dependent variable)	d (Power of Food Scale → dependent variable)	e (Condition → dependent variable)	Indirect effect through mindfulness (a^*c)	Indirect effect through Power of Food Scale (b^*d)	Total effect
Primary outcomes at 1-month follow-up[a]								
Three-Factor Eating Questionnaire-Cognitive Restraint	0.100 (0.046), *P*=.03	–0.093 (0.037), *P*=.01	0.096 (0.055), *P*=.08	0.083 (0.056), *P*=.14	0.006 (0.053), *P*=.90	0.010 (0.007), *P*=.19	–0.008 (0.006), *P*=.21	0.008 (0.051), *P*=.87
Three-Factor Eating Questionnaire-Uncontrolled Eating	N/A[b]	N/A	–0.120 (0.042), *P*=.005	0.447 (0.048), *P*<.001	–0.048 (0.039), *P*=.22	–0.012 (0.007), *P*=.09	–0.042 (0.017), *P*=.02	–0.101 (0.042), *P*=.02
Three-Factor Eating Questionnaire-Emotional Eating	N/A	N/A	–0.049 (0.051), *P*=.33	0.275 (0.054), *P*<.001	–0.103 (0.046), *P*=.03	–0.005 (0.006), *P*=.40	–0.026 (0.011), *P*=.02	–0.134 (0.046), *P*=.004
Primary outcomes at 3-month follow-up[c]								
Three-Factor Eating Questionnaire-Cognitive Restraint	0.100 (0.046), *P*=.03	–0.093 (0.037), *P*=.01	–0.054 (0.055), *P*=.33	0.037 (0.060), *P*=.54	0.029 (0.050), *P*=.57	–0.005 (0.006), *P*=.39	–0.003 (0.006), *P*=.57	0.020 (0.049), *P*=.69
Three-Factor Eating Questionnaire-Uncontrolled Eating	N/A	N/A	–0.176 (0.044), *P*<.001	0.340 (0.053), *P*<.001	–0.071 (0.040), *P*=.08	–0.017 (0.009), *P*=.06	–0.032 (0.013), *P*=.02	–0.121 (0.043), *P*=.005
Three-Factor Eating Questionnaire-Emotional Eating	N/A	N/A	–0.180 (0.046), *P*<.001	0.261 (0.048), *P*<.001	–0.093 (0.041), *P*=.02	–0.018 (0.009), *P*=.06	–0.024 (0.011), *P*=.02	–0.136 (0.043), *P*=.001
Secondary outcomes at 1-month follow-up[d]								
9-item Patient Health Questionnaire	0.100 (0.046), *P*=.03	–0.093 (0.037), *P*=.01	–0.349 (0.053), *P*<.001	0.121 (0.054), *P*=.03	–0.006 (0.048), *P*=.90	–0.035 (0.017), *P*=.04	–0.011 (0.007), *P*=.09	–0.052 (0.049), *P*=.29
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.351 (0.052), *P*<.001	0.080 (0.052), *P*=.12	–0.020 (0.046), *P*=.66	–0.035 (0.017), *P*=.04	–0.007 (0.006), *P*=.19	–0.063 (0.048), *P*=.19
5-item World Health Organization Well-being Index	N/A	N/A	0.206 (0.054), *P*<.001	0.009 (0.052), *P*=.87	0.048 (0.048), *P*=.31	0.021 (0.011), *P*=.07	–0.001 (0.005), *P*=.87	0.068 (0.047), *P*=.15
Secondary outcomes at 3-month follow-up[e]								
9-item Patient Health Questionnaire	0.100 (0.046), *P*=.03	–0.093 (0.037), *P*=.01	–0.337 (0.053), *P*<.001	0.155 (0.051), *P*=.003	–0.028 (0.047), *P*=.545	–0.034 (0.016), *P*=.04	–0.014 (0.008), *P*=.06	–0.077 (0.049), *P*=.12
7-item Generalized Anxiety Disorder Scale	N/A	N/A	–0.211 (0.055), *P*<.001	0.157 (0.054), *P*=.004	–0.031 (0.049), *P*=.52	–0.021 (0.011), *P*=.06	–0.015 (0.008), *P*=.06	–0.067 (0.049), *P*=.18
5-item World Health Organization Well-being Index	N/A	N/A	0.141 (0.060), *P*=.02	–0.034 (0.056), *P*=.54	0.097 (0.052), *P*=.06	0.014 (0.009), *P*=.11	0.003 (0.006), *P*=.56	0.114 (0.051), *P*=.03

[a]Model fit: χ^2_{20}=37.8; *P*=.009; comparative fit index=0.981; Tucker–Lewis index=0.962; standardized root mean squared residual=0.029; root mean square error of approximation=0.050.

[b]N/A: not applicable.

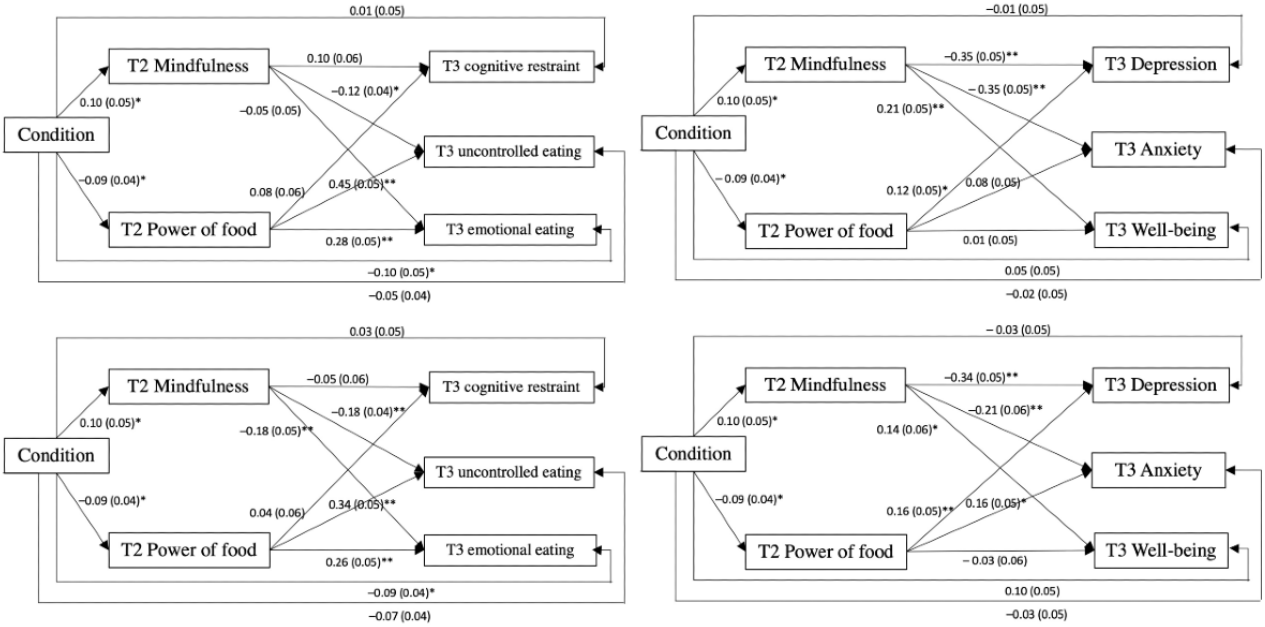
[c]Model fit: χ^2_{20}=32.5; *P*=.04; comparative fit index=0.987; Tucker–Lewis index=0.973; standardized root mean squared residual=0.027; root mean square error of approximation=0.042.

[d]Model fit: χ^2_{20}=44.2; *P*=.002; comparative fit index=0.977; Tucker–Lewis index=0.954; standardized root mean squared residual=0.041; root mean square error of approximation=0.06.

XSL•FO
RenderX

[e]Model fit: χ^2_{20}=28.6; *P*=.09; comparative fit index=0.991; Tucker–Lewis index=0.983; standardized root mean squared residual=0.035; root mean square error of approximation=0.035.

Figure 6. Path models of study 3. Baseline variables were controlled. For the purpose of clarity, the controlled variables and the covariance of the residuals between variables are not depicted. T2: postassessment; T3: 1-month follow-up assessment; T4: 3-month follow-up assessment. *P<.05, **P<.001.



Discussion

Principal Results

This study hypothesized that text messaging–based MBI is effective in reducing distress related to insomnia, pain, and dysfunctional eating. It was encouraging that effectiveness was demonstrated in all 3 randomized controlled trials. Most outcomes had improved significantly at 1-month follow-up compared to wait-list control condition and some effects were able to sustain at 3-month follow-up, including both primary outcomes (eg, insomnia, pain, dysregulated eating indicators) and secondary outcomes (eg, depressive, anxiety symptoms). Our findings showed that text messaging–based interventions not only serve as an adjunct or supplementary tool for traditional treatment programs but also can serve as the core treatment modality. Not only the effectiveness is promising, the retention rates are also high across 3 trials at postmeasurement for intervention groups, ranging from 72.9% (129/177) up to 83.2% (139/167), which demonstrated that text messaging–based intervention is a feasible solution that can maintain a high retention rate. This study contributed to the knowledge of how we can utilize the text messaging application to a fuller extent for intervention purposes so that people with a range of health-related issues can consider evidence-based options delivered via mobile messaging technology.

Study 1 on Insomnia

The results of the randomized controlled trials demonstrated the effectiveness of the text messaging–based MBI on insomnia. Primary outcomes, namely, insomnia intensity, dysfunctional beliefs and attitudes about sleep, and presleep arousal were found to be significantly reduced compared to those in the wait-list control condition, and the effect was sustained at 3-month follow-up. Significant improvement of secondary outcomes, including depressive symptoms, anxiety symptoms, and mental well-being, were also found with sustained effect at 3-month follow-up. The results were consistent with those reported in previous studies on MBIs for insomnia that MBIs are effective in improving insomnia, depression, and anxiety and increasing mindfulness [12,13].

Dysfunctional beliefs and attitudes about sleep mediated the effect of the intervention on all primary outcomes and most secondary outcomes at both 1-month and 3-month follow-up. Mindfulness was found to mediate the effect of the intervention on presleep arousal (cognitive) at 1-month follow-up with additional mediating effects on presleep arousal (somatic) at 3-month follow-up. The mediating effects of dysfunctional beliefs and attitudes about sleep on the outcomes aligned with the mechanism proposed that it was suggested that mindfulness can improve sleep by changing the pattern of worry and rumination [24]. In this study, the mediation model showed that improved sleep might be made possible through reduced dysfunctional beliefs about sleep. However, it was unclear why mindfulness has a mediation effect only on the primary outcomes but not on the secondary outcomes. This pattern might be explained by the specificity of content that was targeted at insomnia.

Study 2 on Pain

The results of repeated measures ANOVA showed that MBIs are effective in enhancing chronic pain acceptance and reducing pain catastrophizing at 3-month follow-up in the intervention group compared to those in the wait-list control group. Mindfulness, depressive symptoms, and mental well-being were

XSL•FO
RenderX

also found to improve significantly at 3-month follow-up. However, no significant improvement was found in pain intensity, which can be explained by the non–symptom-focused approach of acceptance as it aims to improve pain adjustment independent of pain intensity [58]. Other results were largely consistent with previous research that MBI improved pain acceptance, pain catastrophizing, mindfulness, depression, and well-being [14-16]. However, anxiety was not reduced significantly, which might be explained by the floor effect that the baseline score was low. For the mediation model, preliminary evidence suggested that pain catastrophizing mediated the effect of MBI on pain intensity and functioning at both 1 month and 3 months, whereas mindfulness was only found to be mediating the effect of intervention on secondary outcomes but not primary outcomes. Our findings were consistent with a previous study conducted by Elvery et al [25] that pain catastrophizing emerged as the most robust process and it was the most predictive of pain intensity compared to pain acceptance and mindfulness. However, other studies showed a different process and mechanism wherein pain catastrophizing did not mediate the outcome but pain acceptance did, which might imply that the psychological process may be more important than cognitive process [80]. Further research is needed to verify the mechanism of how MBI leads to change in various outcomes.

Study 3 on Dysregulated Eating

MBI was found to be effective in improving uncontrolled eating, emotional eating, and well-being at 3-month follow-up compared to the wait-list control condition. Nonsignificant results were found in other primary and secondary outcomes. However, for data without 3-month follow-up, a significant interaction effect was found for mindfulness. Owing to the complexity of eating issues, the materials used in this condition may not be focused enough to induce significant and sustained change. Preliminary moderation analysis showed that participants with higher uncontrolled eating scores showed greater reduction in uncontrolled eating at postintervention in this study. This finding implies that the intervention may have greater benefits for people who have more severe uncontrolled eating problems. The results of the mediation analysis showed that the power of food mediated the effect of intervention on both uncontrolled and emotional eating at both 1-month and 3-month follow-ups. Mindfulness was also found to mediate the effect on depressive symptoms at both 1-month and 3-month follow-ups. The pattern of mediation is similar to that in the pain condition that mindfulness only mediates secondary outcomes but not primary outcomes. Given little research has been conducted on the relationship between MBI and eating concerns, further research is needed to confirm the model.

Implications

This study took a significant step in demonstrating the effectiveness of using text messaging–based MBIs to alleviate distress related to insomnia, pain, and dysregulated eating, which were affecting at least a quarter of the general population. Such interventions are potentially scalable to the population level and can be widely disseminated with relatively low costs and human resources. This novel way of delivering MBIs showed a high retention rate ranging from 72.9% (129/177) to 83.2% (139/167) for intervention groups at postintervention, which is higher than the median retention rate of internet-based interventions [29]. Given the potency of the text messaging MBIs in alleviating specific health concerns and general psychological distress, text messaging–based intervention deserves more attention in the future in delivering other forms of psychological interventions. Not only as adjunct to traditional face-to-face psychological interventions, text-messaging interventions can potentially become a core intervention modality that has high user engagement and positive treatment outcomes. This study also shed light on the mediation models of MBIs, which have not been thoroughly investigated. Across the 3 randomized controlled trials, mindfulness was found to mediate the effect of intervention on both primary and secondary outcomes. Condition-specific mediators were also found, including dysfunctional beliefs and attitudes about sleep for insomnia, pain catastrophizing for pain, and power of food for dysregulated eating. These findings supported that the cultivation of mindfulness and the ability to reperceive the present experience can alleviate health-related concerns and distress. Future designs for MBIs can consider focusing on these mechanisms of change. Nonetheless, replication is needed to confirm the mediation processes and how different mechanisms are related to different outcomes.

Limitations

This study had several limitations that warrant attention. First, wait-list control was used instead of active control that can account for demand characteristics. As this study focused on investigating the effects of text messaging–based interventions among urban dwellers, comparison with wait-list controls is closer to the real-life situation where people generally do not seek help for their health-related conditions. Even though the placebo effect cannot be ruled out, as the active control group was not included [81], these findings demonstrated effectiveness in a wide range of psychological outcomes for 3 common health concerns.

Second, this study solely adopted self-report measures that may lack objectivity on the severity of health concerns such as sleep habits, pain severity, and eating patterns. Given the interventions were delivered over WhatsApp, the use of self-report measures that are completed online is consistent with accessing the intervention materials over mobile text messaging. Nevertheless, future studies may consider including behavioral or physiological measures to corroborate with self-report findings.

Third, the preponderance of the participants were women. The results of this study may not be generalizable to other genders. The skewness is consistent with gender difference in help-seeking, where men are less likely to seek help than women [82] and women are more interested in practicing mindfulness [83]. A systematic review also found that among 117 studies of randomized controlled trials of mindfulness-based cognitive therapy or mindfulness-based stress reduction with 9820 participants, only 29% of the total participants were men [84]. Future studies on smartphone-based MBIs should sample more men to examine whether these interventions may be more acceptable to men and conducive to their well-being.

XSL•FO
RenderX

Lastly, despite the high retention rates at postmeasurement, the return rates of questionnaires were low at 3-month follow-up, ranging from 61% (72/118) to 67.1% (112/167) for intervention groups and 68.7% (114/166) to 70.1% (82/117) for wait-list control groups. In future studies, strategies are needed to boost the return rates of questionnaires so that the sustained effectiveness can be captured more accurately. One of the possible solutions may be increasing the incentive for completing questionnaires at long follow-up periods.

Conclusions

To conclude, this study showed that text messaging–delivered MBIs are effective in improving issues related to sleep, pain, and eating. Text messaging has the potential to be a core intervention modality to cater to the needs of people with a fast-paced lifestyle or increase accessibility to MBIs. The demonstrated mechanisms of change illuminate directions for future design of materials and focus on MBIs. Given the increasing health needs of the general population and low availability of evidence-based face-to-face interventions, text messaging–based interventions provide a viable alternative to expand on the availability of effective interventions to the public.

Acknowledgments

We would like to express our deepest gratitude to Sania SW Yau, late Chief Executive Officer of New Life Psychiatric Rehabilitation Association, for her vision, leadership, and guidance in driving innovative ways to promote mental health in the community. Even though Sania has left us, she is always in our hearts.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Outline of the intervention for study 1 and study 2.
[PNG File , 355 KB - jmir_v24i5e30073_app1.png]

Multimedia Appendix 2
Outline of the intervention for study 3.
[PNG File , 357 KB - jmir_v24i5e30073_app2.png]

Multimedia Appendix 3
Screenshots of the intervention (translated mock-up).
[PNG File , 2070 KB - jmir_v24i5e30073_app3.png]

Multimedia Appendix 4
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1221 KB - jmir_v24i5e30073_app4.pdf]

References

1. Lam LC, Wong CS, Wang M, Chan W, Chen EY, Ng RM, et al. Prevalence, psychosocial correlates and service utilization of depressive and anxiety disorders in Hong Kong: the Hong Kong Mental Morbidity Survey (HKMMS). Soc Psychiatry Psychiatr Epidemiol 2015 Sep;50(9):1379-1388. [doi: 10.1007/s00127-015-1014-5] [Medline: 25660760]
2. Morin CM, Rodrigue S, Ivers H. Role of stress, arousal, and coping skills in primary insomnia. Psychosom Med 2003;65(2):259-267. [doi: 10.1097/01.psy.0000030391.09558.a3] [Medline: 12651993]
3. Davis MC, Zautra AJ, Smith BW. Chronic pain, stress, and the dynamics of affective differentiation. J Pers 2004 Dec;72(6):1133-1159 [FREE Full text] [doi: 10.1111/j.1467-6494.2004.00293.x] [Medline: 15509279]
4. Tryon M, DeCant R, Laugero K. Having your cake and eating it too: A habit of comfort food may link chronic social stress exposure and acute stress-induced cortisol hyporesponsiveness. Physiology & Behavior 2013 Apr;114-115:32-37. [doi: 10.1016/j.physbeh.2013.02.018]
5. Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why sleep matters—the economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Q 2017;6(4):11-11. [Medline: 5627640]
6. Maniadakis N, Gray A. The economic burden of back pain in the UK. Pain 2000;84(1):95-103. [doi: 10.1016/S0304-3959(99)00187-6]
7. Scarborough P, Bhatnagar P, Wickramasinghe KK, Allender S, Foster C, Rayner M. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006-07 NHS costs. J Public Health (Oxf) 2011 Dec;33(4):527-535. [doi: 10.1093/pubmed/fdr033] [Medline: 21562029]

XSL•FO
RenderX

8. Wong W, Fielding R. Prevalence of insomnia among Chinese adults in Hong Kong: a population-based study. J Sleep Res 2011 Mar;20(1 Pt 1):117-126 [FREE Full text] [doi: 10.1111/j.1365-2869.2010.00822.x] [Medline: 20408932]
9. Wong WS, Fielding R. Prevalence and characteristics of chronic pain in the general population of Hong Kong. J Pain 2011 Feb;12(2):236-245. [doi: 10.1016/j.jpain.2010.07.004] [Medline: 20875775]
10. Centre for health protection, department of health: fruit consumption. Hong Kong Special Administrative Region. URL: https://www.chp.gov.hk/en/statistics/data/10/280/6627.html [accessed 2021-09-01]
11. Kabat-Zinn J. Mindfulness Meditation for Everyday Life. London: Piatkus Books; 2001.
12. Zhang J, Liu X, Xie X, Zhao D, Shan M, Zhang X, et al. Mindfulness-based stress reduction for chronic insomnia in adults older than 75 years: a randomized, controlled, single-blind clinical trial. Explore (NY) 2015;11(3):180-185. [doi: 10.1016/j.explore.2015.02.005] [Medline: 25843539]
13. Klatt M, Norre C, Reader B, Yodice L, White S. Mindfulness in Motion: a Mindfulness-Based Intervention to Reduce Stress and Enhance Quality of Sleep in Scandinavian Employees. Mindfulness 2016 Oct 3;8(2):481-488. [doi: 10.1007/s12671-016-0621-x]
14. Lauche R, Cramer H, Dobos G, Langhorst J, Schmidt S. A systematic review and meta-analysis of mindfulness-based stress reduction for the fibromyalgia syndrome. J Psychosom Res 2013 Dec;75(6):500-510. [doi: 10.1016/j.jpsychores.2013.10.010] [Medline: 24290038]
15. Veehof MM, Oskam M, Schreurs KMG, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542. [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
16. Anheyer D, Haller H, Barth J, Lauche R, Dobos G, Cramer H. Mindfulness-Based Stress Reduction for Treating Low Back Pain: A Systematic Review and Meta-analysis. Ann Intern Med 2017 Jun 06;166(11):799-807 [FREE Full text] [doi: 10.7326/M16-1997] [Medline: 28437793]
17. Godfrey KM, Gallo LC, Afari N. Mindfulness-based interventions for binge eating: a systematic review and meta-analysis. J Behav Med 2015 Apr;38(2):348-362. [doi: 10.1007/s10865-014-9610-5] [Medline: 25417199]
18. O'Reilly GA, Cook L, Spruijt-Metz D, Black DS. Mindfulness-based interventions for obesity-related eating behaviours: a literature review. Obes Rev 2014 Jun;15(6):453-461 [FREE Full text] [doi: 10.1111/obr.12156] [Medline: 24636206]
19. Galante J, Friedrich C, Dawson AF, Modrego-Alarcón M, Gebbing P, Delgado-Suárez I, et al. Mindfulness-based programmes for mental health promotion in adults in nonclinical settings: A systematic review and meta-analysis of randomised controlled trials. PLoS Med 2021 Jan;18(1):e1003481 [FREE Full text] [doi: 10.1371/journal.pmed.1003481] [Medline: 33428616]
20. Querstret D, Morison L, Dickinson S, Cropley M, John M. Mindfulness-based stress reduction and mindfulness-based cognitive therapy for psychological health and well-being in nonclinical samples: A systematic review and meta-analysis. International Journal of Stress Management 2020 Nov;27(4):394-411. [doi: 10.1037/str0000165]
21. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
22. Nyklícek I, Kuijpers KF. Effects of mindfulness-based stress reduction intervention on psychological well-being and quality of life: is increased mindfulness indeed the mechanism? Ann Behav Med 2008 Jun;35(3):331-340 [FREE Full text] [doi: 10.1007/s12160-008-9030-2] [Medline: 18535870]
23. Shapiro SL, Carlson LE, Astin JA, Freedman B. Mechanisms of mindfulness. J Clin Psychol 2006 Mar;62(3):373-386. [doi: 10.1002/jclp.20237] [Medline: 16385481]
24. Lundh L. The Role of Acceptance and Mindfulness in the Treatment of Insomnia. J Cogn Psychother 2005 Mar 01;19(1):29-39. [doi: 10.1891/jcop.19.1.29.66331]
25. Elvery N, Jensen M, Ehde D, Day M. Pain catastrophizing, mindfulness, and pain acceptance: what?s the difference? Clin J Pain 2017;33(6):485-495. [doi: 10.1097/ajp.0000000000000430]
26. Kristeller J, Wolever RQ, Sheets V. Mindfulness-Based Eating Awareness Training (MB-EAT) for Binge Eating: A Randomized Clinical Trial. Mindfulness 2013 Feb 1;5(3):282-297. [doi: 10.1007/s12671-012-0179-1]
27. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The Effectiveness of Prompts to Promote Engagement With Digital Interventions: A Systematic Review. J Med Internet Res 2016 Jan 08;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
28. 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;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
29. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
30. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]

XSL•FO
RenderX

31. Barak A, Grohol JM. Current and Future Trends in Internet-Supported Mental Health Interventions. Journal of Technology in Human Services 2011 Jul;29(3):155-196. [doi: 10.1080/15228835.2011.616939]
32. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
33. Berrouiguet 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 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
34. Shaw R, Bosworth H. Short message service (SMS) text messaging as an intervention medium for weight loss: A literature review. Health Informatics J 2012 Dec;18(4):235-250 [FREE Full text] [doi: 10.1177/1460458212442422] [Medline: 23257055]
35. Sibeko G, Temmingh H, Mall S, Williams-Ashman P, Thornicroft G, Susser ES, et al. Improving adherence in mental health service users with severe mental illness in South Africa: a pilot randomized controlled trial of a treatment partner and text message intervention vs. treatment as usual. BMC Res Notes 2017 Nov 09;10(1):584 [FREE Full text] [doi: 10.1186/s13104-017-2915-z] [Medline: 29121999]
36. Sims H, Sanghara H, Hayes D, Wandiembe S, Finch M, Jakobsen H, et al. Text message reminders of appointments: a pilot intervention at four community mental health clinics in London. Psychiatr Serv 2012 Feb 01;63(2):161-168. [doi: 10.1176/appi.ps.201100211] [Medline: 22302334]
37. Lee HY, Koopmeiners JS, Rhee TG, Raveis VH, Ahluwalia JS. Mobile phone text messaging intervention for cervical cancer screening: changes in knowledge and behavior pre-post intervention. J Med Internet Res 2014 Aug 27;16(8):e196 [FREE Full text] [doi: 10.2196/jmir.3576] [Medline: 25164545]
38. Pereira AAC, Destro JR, Picinin Bernuci M, Garcia LF, Rodrigues Lucena TF. Effects of a WhatsApp-Delivered Education Intervention to Enhance Breast Cancer Knowledge in Women: Mixed-Methods Study. JMIR Mhealth Uhealth 2020 Jul 21;8(7):e17430 [FREE Full text] [doi: 10.2196/17430] [Medline: 32706726]
39. Whatsapp Inc. URL: https://whatsapp.com [accessed 2021-09-01]
40. Most popular messaging apps. Statista. URL: https://www.statista.com/statistics/258749/most-popular-global-mobile-messenger-apps/ [accessed 2020-12-13]
41. Microsoft excel. Microsoft Corporation. 2010. URL: https://office.microsoft.com/excel [accessed 2021-09-10]
42. Qualtrics XM. 2017. URL: https://www.qualtrics.com [accessed 2021-09-10]
43. Hong YW. Efficacy of a Valued Living Program Delivered Through Instant Messaging Application in Promoting Mental Well-being of College Students in Hong Kong [master thesis]. Hong Kong: The Chinese University of Hong Kong; 2017.
44. Kabat-Zinn J. Full Catastrophe Living, Revised Edition: How to Cope with Stress, Pain and Illness Using Mindfulness Meditation. Hachette UK: Little, Brown Book Group; 2013.
45. Bastien C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Medicine 2001;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4]
46. Cincotta AL, Gehrman P, Gooneratne NS, Baime MJ. The effects of a mindfulness-based stress reduction programme on pre-sleep cognitive arousal and insomnia symptoms: a pilot study. Stress and Health 2010 Dec 19;27(3):e299-e305. [doi: 10.1002/smi.1370]
47. Gross CR, Kreitzer MJ, Reilly-Spong M, Wall M, Winbush NY, Patterson R, et al. Mindfulness-based stress reduction versus pharmacotherapy for chronic primary insomnia: a randomized controlled clinical trial. Explore (NY) 2011;7(2):76-87 [FREE Full text] [doi: 10.1016/j.explore.2010.12.003] [Medline: 21397868]
48. Ong JC, Shapiro SL, Manber R. Mindfulness meditation and cognitive behavioral therapy for insomnia: a naturalistic 12-month follow-up. Explore (NY) 2009;5(1):30-36 [FREE Full text] [doi: 10.1016/j.explore.2008.10.004] [Medline: 19114261]
49. Kan KK. Validation of the insomnia severity index, Athens insomnia scale and sleep quality index in adolescent population in Hong Kong. Published online 2008, b4073 2008:3634. [doi: 10.5353/th_b4073363]
50. Chung KF, Kan KKK, Yeung WF. Assessing insomnia in adolescents: comparison of Insomnia Severity Index, Athens Insomnia Scale and Sleep Quality Index. Sleep Med 2011 May;12(5):463-470. [doi: 10.1016/j.sleep.2010.09.019] [Medline: 21493134]
51. Nicassio PM, Mendlowitz DR, Fussell JJ, Petras L. The phenomenology of the pre-sleep state: The development of the pre-sleep arousal scale. Behaviour Research and Therapy 1985;23(3):263-271. [doi: 10.1016/0005-7967(85)90004-x]
52. Brislin RW. The wording and translation of research instruments. In: Field Methods in Cross-Cultural Research. Canada: Sage Publications, Inc; 1986:137-164.
53. Morin C, Vallières A, Ivers H. Dysfunctional beliefs and attitudes about sleep (DBAS): validation of a brief version (DBAS-16). Sleep 2007 Nov;30(11):1547-1554 [FREE Full text] [doi: 10.1093/sleep/30.11.1547] [Medline: 18041487]
54. Chen CW, Jan YW, Yang CM, Lin SC. Dysfunctional beliefs and attitudes about sleep (DBAS): validation of the Chinese version. Archives of Clinical Psychology. 2009. URL: https://tinyurl.com/3eafffhc [accessed 2022-04-22]
55. Kelly AM. The minimum clinically significant difference in visual analogue scale pain score does not differ with severity of pain. Emerg Med J 2001 May;18(3):205-207 [FREE Full text] [doi: 10.1136/emj.18.3.205] [Medline: 11354213]

XSL•FO
RenderX

56. Herr KA, Spratt K, Mobily PR, Richardson G. Pain intensity assessment in older adults: use of experimental pain to compare psychometric properties and usability of selected pain scales with younger adults. Clin J Pain 2004;20(4):207-219. [doi: 10.1097/00002508-200407000-00002] [Medline: 15218405]
57. Joyce CRB, Zutshi DW, Hrubes V, Mason RM. Comparison of fixed interval and visual analogue scales for rating chronic pain. Eur J Clin Pharmacol 1975;8(6):415-420. [doi: 10.1007/bf00562315]
58. McCracken LM. Learning to live with the pain: acceptance of pain predicts adjustment in persons with chronic pain. Pain 1998;74(1):21-27. [doi: 10.1016/s0304-3959(97)00146-2]
59. Cheung MN, Ning MC, Wong TCM, Ming TWC, Yap JCM, Mae JYC, et al. Validation of the Chronic Pain Acceptance Questionnaire (CPAQ) in Cantonese-speaking Chinese patients. J Pain 2008 Sep;9(9):823-832. [doi: 10.1016/j.jpain.2008.04.005] [Medline: 18632309]
60. Sullivan MJL, Bishop SR, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
61. Yap JC, Lau J, Chen PP, Gin T, Wong T, Chan I, et al. Validation of the Chinese Pain Catastrophizing Scale (HK-PCS) in patients with chronic pain. Pain Med 2008 Mar 01;9(2):186-195. [doi: 10.1111/j.1526-4637.2007.00307.x] [Medline: 18298701]
62. Karlsson J, Persson L, Sjöström L, Sullivan M. Psychometric properties and factor structure of the Three-Factor Eating Questionnaire (TFEQ) in obese men and women. Results from the Swedish Obese Subjects (SOS) study. Int J Obes Relat Metab Disord 2000 Dec;24(12):1715-1725. [doi: 10.1038/sj.ijo.0801442] [Medline: 11126230]
63. Cappelleri JC, Bushmakin AG, Gerber RA, Leidy NK, Sexton CC, Karlsson J, et al. Evaluating the Power of Food Scale in obese subjects and a general sample of individuals: development and measurement properties. Int J Obes (Lond) 2009 Aug;33(8):913-922. [doi: 10.1038/ijo.2009.107] [Medline: 19506564]
64. Lowe MR, Butryn ML, Didie ER, Annunziato RA, Thomas JG, Crerand CE, et al. The Power of Food Scale. A new measure of the psychological influence of the food environment. Appetite 2009 Aug;53(1):114-118. [doi: 10.1016/j.appet.2009.05.016] [Medline: 19500623]
65. 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] [Medline: 12703651]
66. Carlson LE, Brown KW. Validation of the Mindful Attention Awareness Scale in a cancer population. J Psychosom Res 2005 Jan;58(1):29-33. [doi: 10.1016/j.jpsychores.2004.04.366] [Medline: 15771867]
67. Dharma DIOLA. ZEN – research on creative meditation space for ease and mindfulness research (report no. nsc 99-2218-e-655-001-). Dharma Drum Institute of Liberal Arts. URL: http://www.etop.org.tw/index.php?c=adm11252&m=getReportFile&d=adm&i=135674 [accessed 2022-04-11]
68. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
69. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
70. Wellbeing measures in primary health care: the DepCare project. World Health Organization. URL: https://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf [accessed 2022-04-11]
71. Henkel V, Mergl R, Kohnen R, Maier W, Möller HJ, Hegerl U. Identifying depression in primary care: a comparison of different methods in a prospective cohort study. BMJ 2003 Jan 25;326(7382):200-201 [FREE Full text] [doi: 10.1136/bmj.326.7382.200] [Medline: 12543837]
72. Newnham EA, Hooke GR, Page AC. Monitoring treatment response and outcomes using the World Health Organization's Wellbeing Index in psychiatric care. J Affect Disord 2010 Apr;122(1-2):133-138. [doi: 10.1016/j.jad.2009.06.005] [Medline: 19592116]
73. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
74. Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. Br J Psychiatry 2002 May;180:461-464. [doi: 10.1192/bjp.180.5.461] [Medline: 11983645]
75. Thandi G, Fear NT, Chalder T. A comparison of the Work and Social Adjustment Scale (WSAS) across different patient populations using Rasch analysis and exploratory factor analysis. J Psychosom Res 2017 Jan;92:45-48. [doi: 10.1016/j.jpsychores.2016.11.009] [Medline: 27998511]
76. van Ginkel JR, Kroonenberg PM. Analysis of Variance of Multiply Imputed Data. Multivariate Behav Res 2014 Jan 01;49(1):78-91 [FREE Full text] [doi: 10.1080/00273171.2013.855890] [Medline: 24860197]
77. Rubin D. Multiple Imputation for Nonresponse in Surveys. Wiley 1987:223-243. [doi: 10.1002/9780470316696]
78. Van Ginkel JR. Investigation of Multiple Imputation in Low-Quality Questionnaire Data. Multivariate Behav Res 2010 May 28;45(3):574-598. [doi: 10.1080/00273171.2010.483373] [Medline: 26760493]
79. Muthen L, Muthen B. Mplus Version 7 User's Guide. 2012. URL: https://www.statmodel.com/download/usersguide/Mplus%20user%20guide%20Ver_7_r6_web.pdf [accessed 2022-04-15]
80. Day MA, Thorn BE. The mediating role of pain acceptance during mindfulness-based cognitive therapy for headache. Complement Ther Med 2016 Apr;25:51-54. [doi: 10.1016/j.ctim.2016.01.002] [Medline: 27062948]

XSL•FO
RenderX

81. Vickers A, de CA. Why use placebos in clinical trials? A narrative review of the methodological literature. J Clin Epidemiol 2000;53(2):157-161. [doi: 10.1016/S0895-4356(99)00139-0]
82. Liddon L, Kingerlee R, Barry JA. Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. Br J Clin Psychol 2018 Mar;57(1):42-58. [doi: 10.1111/bjc.12147] [Medline: 28691375]
83. Carlson LE. Uptake of mindfulness - based interventions: A phenomenon of wealthy white western women? Clinical Psychology: Science and Practice 2018 Sep;25(3):e12258. [doi: 10.1111/cpsp.12258]
84. Bodenlos JS, Strang K, Gray-Bauer R, Faherty A, Ashdown BK. Male Representation in Randomized Clinical Trials of Mindfulness-Based Therapies. Mindfulness 2016 Nov 28;8(2):259-265. [doi: 10.1007/s12671-016-0646-1]

Abbreviations

ANOVA: analysis of variance
CONSORT: Consolidated Standards of Reporting Trials
MBI: mindfulness-based intervention
TFEQ-R18: 18-item Three-Factor Eating Questionnaire-Revised
VAS: visual analogue scale

Edited by A Mavragani; submitted 06.05.21; peer-reviewed by D Saleh, N Khalili-Mahani; comments to author 20.09.21; revised version received 09.02.22; accepted 24.02.22; published 03.05.22.

Please cite as:
Li ACM, Wong KKL, Chio FHN, Mak WWS, Poon LWH
Delivering Mindfulness-Based Interventions for Insomnia, Pain, and Dysfunctional Eating Through a Text Messaging App: Three Randomized Controlled Trials Investigating the Effectiveness and Mediating Mechanisms
J Med Internet Res 2022;24(5):e30073
URL: https://www.jmir.org/2022/5/e30073
doi:10.2196/30073
PMID:35503653

©Amanda CM Li, Keith KL Wong, Floria HN Chio, Winnie WS Mak, Loretta WH Poon. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Effects of a Web-Based Tool for Parents of Children With Juvenile Idiopathic Arthritis: Randomized Controlled Trial

Kathleen Mulligan[1,2*], BSc, MSc, PhD; Shashivadan P Hirani[1*], BSc, MSc, PhD; Sally Harris[3], BSc, MSc, ClinPsyD, PhD; Jo Taylor[1], BSc, MSc; Lucy R Wedderburn[4,5,6], BA, MBBS, PhD; Stanton Newman[1], Dip Psych, DPhil; WebParC Investigator group[7]

[1]Centre for Health Services Research, School of Health Sciences, City, University of London, London, United Kingdom

[2]East London NHS Foundation Trust, London, United Kingdom

[3]Brighton and Sussex University Hospitals NHS Trust, Brighton, United Kingdom

[4]University College London Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

[5]Centre for Adolescent Rheumatology Versus Arthritis, London, United Kingdom

[6]National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital, London, United Kingdom

[7]See Acknowledgments

[*]these authors contributed equally

Corresponding Author:
Stanton Newman, Dip Psych, DPhil
Centre for Health Services Research, School of Health Sciences
City, University of London
Northampton Square
London, EC1V 0HB
United Kingdom
Phone: 44 020 7040 3090
Email: Stanton.Newman.1@city.ac.uk

Abstract

Background: Juvenile idiopathic arthritis (JIA) is a group of autoinflammatory diseases that cause pain and disability if not controlled by treatment. Parenting a child with JIA is stressful for parents, who express concerns about their child's treatment and may experience anxiety and powerlessness concerning their child's illness. Parenting stress is greater in parents of children with chronic illness than in those with healthy children and is related to poorer psychological adjustment in both parents and children. It is therefore important to develop interventions to support parents. This paper reports the evaluation of a web-based tool that provides information and practical skills to help increase parents' confidence in managing their child's illness and reduce parenting stress.

Objective: The aim of this study is to evaluate the benefits of a web-based tool (*WebParC*) for parents of children with recently diagnosed JIA.

Methods: A multicentered randomized controlled trial was conducted at pediatric rheumatology centers in England. We recruited parents of children aged ≤12 years who had been diagnosed with JIA within the previous 6 months. They were randomized to the intervention (WebParC access plus standard care) or the control (standard care alone) and followed up 4 months and 12 months after randomization. Where both parents participated, they were randomized *by household* to the same trial arm. The WebParC intervention consists of information about JIA and its treatment plus a toolkit, based on cognitive behavioral therapy, to help parents develop skills to manage JIA-related issues. The primary outcome was the self-report Pediatric Inventory for Parents measure of illness-related parenting stress. The secondary outcomes were parental mood, self-efficacy, coping, effectiveness of participation in their child's health care, satisfaction with health care, and child's health-related quality of life.

Results: A total of 203 *households* comprising 220 parents were randomized to the intervention (100/203, 49.3%) or control (103/203, 50.7%) arm. Follow-up assessments were completed by 65.5% (133/203) of the households at 4 months (intervention 60/100, 60%, and control 73/103, 70.9%) and 61.1% (124/203) of the households at 12 months (intervention 58/100, 58%, and control 66/103, 64.1%). A main effect of the trial arm was found on the Pediatric Inventory for Parents: the intervention participants reported less frequency (subscales *communication* $F_{1,120627}$=5.37; P=.02, and *role function* $F_{1,27203}$=5.40; P=.02) and difficulty

XSL•FO
RenderX

(subscales *communication* $F_{1,2237}$=7.43; *P*=.006, *medical care* $F_{1,2907}$=4.04; *P*=.04, and *role function* $F_{1,821}$=4.37, *P*=.04) regarding illness-related stressful events than the control participants.

Conclusions: The WebParC website for parents of children with JIA reduced illness-related parenting stress. This web-based intervention offers a feasible preventive approach for parents of children with JIA and potentially could be adapted and evaluated for parents of children with other chronic illnesses.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN) 13159730; http://www.isrctn.com/ISRCTN13159730

(J Med Internet Res 2022;24(5):e29787) doi:10.2196/29787

KEYWORDS

parenting stress; juvenile idiopathic arthritis; web-based intervention; randomized controlled trial; parenting; pediatrics; arthritis; RCT; rheumatology; children; youth; web-based tool; mobile phone

Introduction

Background

Parenting a child with juvenile idiopathic arthritis (JIA) presents many challenges, including dealing with the child's pain, distress, and physical difficulties; managing medication, hospital visits (which may involve traveling a considerable distance from home), and impact on schooling; financial issues such as time off work; and uncertainty about the future. In addition, in some health care systems, there are substantial medication costs. Parents of children with JIA have concerns about their child's treatment [1-3] and may experience anxiety and powerlessness concerning their child's illness [4]. Parenting stress is greater in parents of children with chronic illness than in those with healthy children [5] and is associated with poorer psychological adjustment in both parents and children [5,6]. Given the interconnectedness between parent and child adjustment, early intervention to support parents may facilitate better adjustment for their children with JIA [7,8]. The Pediatric Psychosocial Preventative Health Model developed by Kazak [8] is a 3-tier model for treating the families of children in pediatric health settings. The model offers a guide for matching psychosocial support to families' level of need. It proposes that most families of children with chronic illnesses are likely to be distressed but resilient (universal tier). A smaller group of families have risk factors for ongoing difficulties and require targeted care. The smallest group, clinical/treatment, has a high level of risk factors for ongoing distress and requires more intensive clinical services. Kazak [8] stresses the need to adopt preventive approaches to support families in the universal tier to build their resilience and prevent future problems.

A potential preventive approach is to provide web-based interventions. It is important that parents are able to access information from a trusted source [9]; however, health information on the internet is unregulated, often not validated through a systematic process [10], and the quality is variable [11,12]. When developing this research, none of the websites providing information for children and young people with JIA and their parents had been evaluated in a randomized controlled trial (RCT) and none provided skills training in techniques to help parents to manage their child's arthritis [12].

We developed a website for parents of children with recently diagnosed JIA (*WebParC*) [13] to complement usual clinical care, with potential to help parents cope with the stress of their child having JIA. It is a specially designed web-based tool providing around-the-clock access to information and practical skills in dealing with specific problems (eg, taking medication) and accessible as need arises. This paper reports the evaluation of WebParC.

Objectives

The aim of the study is to test the hypothesis that parents provided with WebParC in addition to standard care would experience less illness-related parenting stress than those provided standard care alone.

Methods

Design

This was a multicenter RCT.

Participating Research Sites

A total of 16 National Health Service tertiary pediatric rheumatology services in England participated in the study.

Ethical Approval

Approval was obtained from the Health Research Authority London Bridge Research Ethics Committee (13/LO/0288).

Research Participants

The participants were parents attending rheumatology clinic appointments with their child, who met the criteria outlined in Textbox 1.

XSL•FO
RenderX

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Parent aged ≥18 years, with a child aged ≤12 years, recently diagnosed with juvenile idiopathic arthritis (within the previous 6 months)
  - It was considered appropriate to focus the website on this age group because responsibility is taken mainly by parents in the child's earlier years but tends to move to the child as they get older; therefore, different strategies may be required for parents of adolescents
  - Juvenile idiopathic arthritis was diagnosed by a pediatric rheumatologist according to current International League of Associations for Rheumatology criteria [14], which specify that juvenile idiopathic arthritis involves inflammation of the joints that begins before the age of 16 years and persists for at least 6 weeks. The International League of Associations for Rheumatology categorizes 7 juvenile idiopathic arthritis subtypes that differ in clinical course and are based on the number of inflamed joints, laboratory tests, and clinical features. The subtypes are oligoarticular, polyarticular–rheumatoid factor negative, polyarticular–rheumatoid factor positive, systemic-onset, psoriatic, enthesitis-related, and undifferentiated arthritis
  - One or both parents could participate. Parents did not need to be living together or with the child with juvenile idiopathic arthritis to be eligible
- Internet access
- Able to speak and read English

Exclusion criteria

- Current severe mental illness such as identifiable psychosis in parents
- Major problems with literacy, making the questionnaire completion impossible
- Likely to be distressed by the study, as judged by their child's rheumatologist

Procedures

Overview

Parents were invited to participate by their child's rheumatologist when attending a clinic appointment with their child. Interested parents were given the opportunity to ask questions and were given the participant information sheet to read at the clinic or to take home if they wished to have more time to consider participation. Those who wished to participate provided written consent to the site research staff. If the child with JIA was aged 8-12 years, the child's assent was obtained for their demographic and clinical data to be collected for the research. After providing consent, participants were given baseline questionnaires to complete at the clinic or at home and return them to the trial coordinating center (a freepost envelope was provided). A link to a web-based version of the questionnaire on Qualtrics software was also provided so that parents could choose their preferred completion mode. Where both parents participated, they were given questionnaire packs with separate return envelopes and individual study IDs that they entered into Qualtrics if they chose web-based completion. If the baseline questionnaire was not returned, a member of the site research team contacted the participants by telephone. This was a change to the protocol made partway through the trial to enhance questionnaire return rates.

Randomization

To minimize selection bias, participants were randomized by the trial coordinating center after receipt of the completed baseline questionnaire. Randomization was in a ratio of 1:1 to trial arms. Where both parents participated, they were randomized to the same trial arm (ie, randomization was clustered by *household*). Blocked randomization per site was performed using computer-generated randomization sequences that allowed varying randomization block sizes. A combination of block sizes was used, varying among 2, 4, and 6, depending on site size; we used different-sized blocks so that sites could not guess which group the last participant of a block would be randomized to. Allocation was concealed from clinical teams to avoid biasing clinical care; however, after allocation, it was not possible to blind the trial coordinator because the follow-up questionnaires contained additional questions about the website for intervention arm participants. Other members of the investigating team were blinded to trial arm allocation. Participants were requested not to inform their child's clinicians of their trial arm allocation.

Trial Arms

Control Arm

Children of control arm participants continued to receive standard clinical care as provided by the study site.

Intervention Arm

In addition to standard care, those allocated to the intervention arm were given free unlimited password-protected access to the website.

Website Design

The *JIA website for Parents* site was designed following:

1. A review was conducted of the literature on parents' experiences of having a child with JIA.
2. A review of websites was conducted to find those that (1) included information about JIA for parents, (2) provided specific skills training for parents to manage their child's JIA, and (3) contained information in English. Although many sites were found that provided information about JIA, at the time of review, 5 main websites [15-19] that contained

XSL•FO
RenderX

significant information for parents were reviewed in detail but none contained skills training to assist parents.
3. A focus group was conducted with 6 parents to ask their views on what the website should include.
4. We conducted 2 focus groups with 12 health care professionals—6 (50%) rheumatologists, 5 (42%) rheumatology nurse specialists, and 1 (8%) clinical psychologist—to ask their views on what the website should include.

Website content was written by health professionals supported by a research assistant. The health professionals included 13 rheumatologists, 4 rheumatology nurse specialists, 2 clinical psychologists, an occupational therapist, an ophthalmologist, 2 physiotherapists, a podiatrist, and a social worker. A website consultant designed the site for layout, usability, and acceptability.

The resulting prototype website was tested by 7 parents and eight health professionals (4, 50%, rheumatologists; 2, 25%, rheumatology nurse specialists; 1, 13%, physiotherapist; and 1, 13%, clinical psychologist) to evaluate usability, navigation, structure, layout, and content. Minor changes were made to the website after this assessment. These included condensing some of the text, improving some text formatting with the use of bullet points, correcting a few navigation links, and renaming some tabs with more user-friendly terms.

Over the course of the website development but before trial commencement, the website was reviewed and updated to ensure that user interfaces and content were current. Website content did not change thereafter during the trial period. The website is device adaptive; therefore, it is suitable for use on computer, tablet, and smartphone.

The website has two main components:

1. Information about JIA and its treatment. This comprises sections about cause, diagnosis, JIA types and symptoms, how JIA changes with time, possible complications, the rheumatology team, and everyday life and available treatments. It also includes videos of health professionals explaining JIA and its treatment as well as video testimonials from parents about living with, and caring for, children with JIA as a family.
2. A JIA toolkit based on cognitive behavioral therapy that includes psychoeducation about thoughts, feelings, and behavior following a diagnosis; cognitive restructuring techniques to challenge unhelpful thinking to promote coping with JIA; problem-solving skills to promote coping with adherence issues and stressful events; strategies to promote effective communication with family members and the health care team; and pain management techniques, including cognitive restructuring, relaxation, distraction, and pacing.

The toolkit includes a number of downloadable resources such as problem-solving sheet, thought diary, breaking negative thought cycle sheet, reward chart, procedure contract template and certificate, visual timetable, and audio relaxation sessions for children and adults.

Trial Measures

Parent Data

Parents provided demographic data including age, gender, education level, and relationship status.

Information on the validated self-report questionnaire measures is reported in Table 1. The primary outcome was parenting stress at 4 months after randomization, measured with the Pediatric Inventory for Parents (PIP) [20], which is a validated measure to assess difficult events that parents may face. Respondents answer two questions for each event: how often it occurred in the past 7 days and how difficult it was for them. The 4-month time span was chosen to give parents sufficient time to make use of the website and to evaluate its effect in the short term. Follow-up times were also selected to fit around clinic visits.

The secondary outcomes were as follows:

- Parenting stress at 12 months after randomization using the PIP [20]. This time span was chosen to evaluate the medium-term effects of using the website.
- Parent mood, assessed with the Hospital Anxiety and Depression Scale [21].
- Parent confidence in managing their child's arthritis, assessed with the Parent's Arthritis Self-Efficacy Scale (PASE) [22].
- Parent effectiveness in managing their child's health care, assessed using the Effective Consumer Scale–Adapted (ECS17-A) [23]. The original scale developed for adults with musculoskeletal disease was adapted to refer to how parents manage their child's disease.
- Parent satisfaction with health care, assessed with the Client Satisfaction Questionnaire [24].
- A proxy measure of the child's health-related quality of life was assessed with the Child Health Questionnaire, 50-item parent version (CHQ-PF50) [25].

Process measures on website use and parent coping and beliefs about their child's illness (Brief Coping Orientation to Problems Experienced [26] and Brief Illness Perception Questionnaire [27], respectively) were collected but will be reported separately from this paper, which focuses on the trial outcomes.

XSL•FO
RenderX

Table 1. Trial measures.

Measure and subscales	Number of items	Response scale	Scoring	Cronbach α in Web-ParC study at baseline
Pediatric Inventory for Parents [20]		1=never to 5=very often and 1=not at all to 5=extremely	Higher score=greater frequency or difficulty of stressful events	
2 subscales (Frequency and Difficulty)	84 (42 for each subscale)		Total 42-210 in each subscale	
4 domains			Frequency total score: .959; Difficulty total score: .965	
Communication	9		9-45	Frequency: .787; Difficulty: .841
Emotional distress	15		15-75	Frequency: .918; Difficulty: .913
Medical care	8		8-40	Frequency: .840; Difficulty: .846
Role function	10		10-50	Frequency: .840; Difficulty: .864
Hospital Anxiety and Depression Scale [21]				
Anxiety and Depression	7 items per scale	0-3 (response options vary for each item)	0-21 per subscale (mild: 8-10; moderate: 11-14; severe: 15-21); higher score=more symptoms of anxiety or depression	Anxiety: .900; Depression: .872
Parent's Arthritis Self-Efficacy Scale [22]				
Symptoms and Psychosocial	14 (7 in each subscale)	1=very uncertain to 7=very certain or not applicable	Scores are standardized to 0-10 for each subscale. Higher score=greater self-efficacy	Symptoms: .902; Psychosocial: .934
Effective Consumer Scale–Adapted [23]	17	0=never to 4=always	A score for each domain and a total score are calculated and converted to 0-100. Higher score=a more effective consumer of health care	Total score: .933
Use of health information	3			.812
Clarifying personal priorities	3			.836
Communicating with others	3			.801
Negotiating roles and taking control	4			.798
Deciding and taking action	4			.876
Client Satisfaction Questionnaire [24]				
N/A[a]	8	1-4 (response options vary for each item)	A total score is calculated. Higher score=greater satisfaction with healthcare	.898
Child Health Questionnaire 50-item parent version [25]	50	Variable	Standardized to population norms and range from 0 to 100 (mean 50, SD 10). Higher score=better health-related quality of life	
Summary scores				N/A
Physical				
Psychosocial				
Subscales				
Physical functioning				.942
Role/social limitations–emotional/behavioral				.988
Role/social limitations–physical				.962

XSL•FO
RenderX

Measure and subscales	Number of items	Response scale	Scoring	Cronbach α in Web-ParC study at baseline
Bodily pain and discomfort				.923
Behavior				.905
Mental health				.921
Self-esteem				.954
General health perceptions				.709
Parental impact–emotional				.821
Parental impact–time				.807
Family activities				.940
Family cohesion				N/A: single item

[a]N/A: not applicable.

Child Data

Information about participants' children with JIA (gender, age, JIA subtype, date of diagnosis, core outcome variables (number of inflamed and limited joints, erythrocyte sedimentation rate, Child Health Assessment Questionnaire, parent global rating, and physician global rating) [28], medication, and any related comorbidities were gathered by the clinical sites and sent securely to the trial coordinating center.

Follow-up

At 4 and 12 months after randomization, the trial center sent follow-up questionnaires both electronically and in hard copy for participants to choose their preferred completion method. Where both parents participated, they were mailed the follow-up questionnaire packs individually and a link to the web-based questionnaire was sent to their individual email addresses. Up to two telephone or text reminders were sent if questionnaires were not returned within 2 weeks. Follow-up clinical data for the child (core outcome variables, medication, and comorbidities) were obtained from trial sites' clinic notes closest to the follow-up time points.

Participants were sent a £5 (US $6.80) gift voucher on return of each completed study questionnaire. This protocol change was made partway through the trial to enhance questionnaire return rates, but the gift voucher was offered retrospectively to all participants.

Statistical Considerations

Sample Size

Both parents were invited to participate; therefore, sample size calculation allowed for *clustering* by household. The power calculation was based on the PIP primary outcome measure, assessed at 4 months after randomization. SDs on the PIP scales frequency (PIP-F) and difficulty (PIP-D) were expected to be 25 [20]. Therefore, 85 households per trial arm was considered adequate to detect a mean difference of 10 points with 80% power and 5% significance level, representing a medium effect size. This allowed for clustering by household, assuming an intracluster correlation of 0.5. Allowing for a 15% dropout rate, 100 households per trial arm (200 total) were needed.

Statistical Analysis

Data were collected and stored in a secure manner in accordance with the guidelines of the United Kingdom's Data Protection Act and the European Union's General Data Protection Regulation. Analysis was conducted using SPSS software (version 25.0; IBM Corp).

Missing value analysis examined item-level missing data. Scale authors' rules, where available, were applied for dealing with missing data. If rules were not available, mean imputation within a scale was used when ≥50% of the scale items were available. The Little missing completely at random (MCAR) test was conducted to indicate the appropriateness of further imputations. If the levels of missing data on any scale or item were >10%, multiple imputation was conducted (m=10). Data from all time points (baseline, 4 months, and 12 months) were used to predict missing data, but the 3 time points were imputed separately and only for participants who provided data at that time point. Resultant databases were analyzed separately, after which Rubin's rules [29,30] were used to combine the results from the 10 data sets. Responders (those who completed at least one follow-up) were compared with nonresponders on baseline characteristics using logistic regression analyses.

Analyses were on an intention-to-treat basis. Outcomes were compared using multilevel modeling with a random effect of household and adjusting for the variable at baseline and any parent and child demographic characteristics that differed between trial arms. We explored the main effects of time and trial arm and their interaction by entering trial arm, time, and the interaction between trial arm and time as fixed effects, using the restricted maximum likelihood estimation method. Significant interaction terms were interpreted as indicating differential treatment effectiveness and explored with post hoc tests. Hedges *g* was calculated for effect sizes of differences between trial arms at each follow-up.

Results

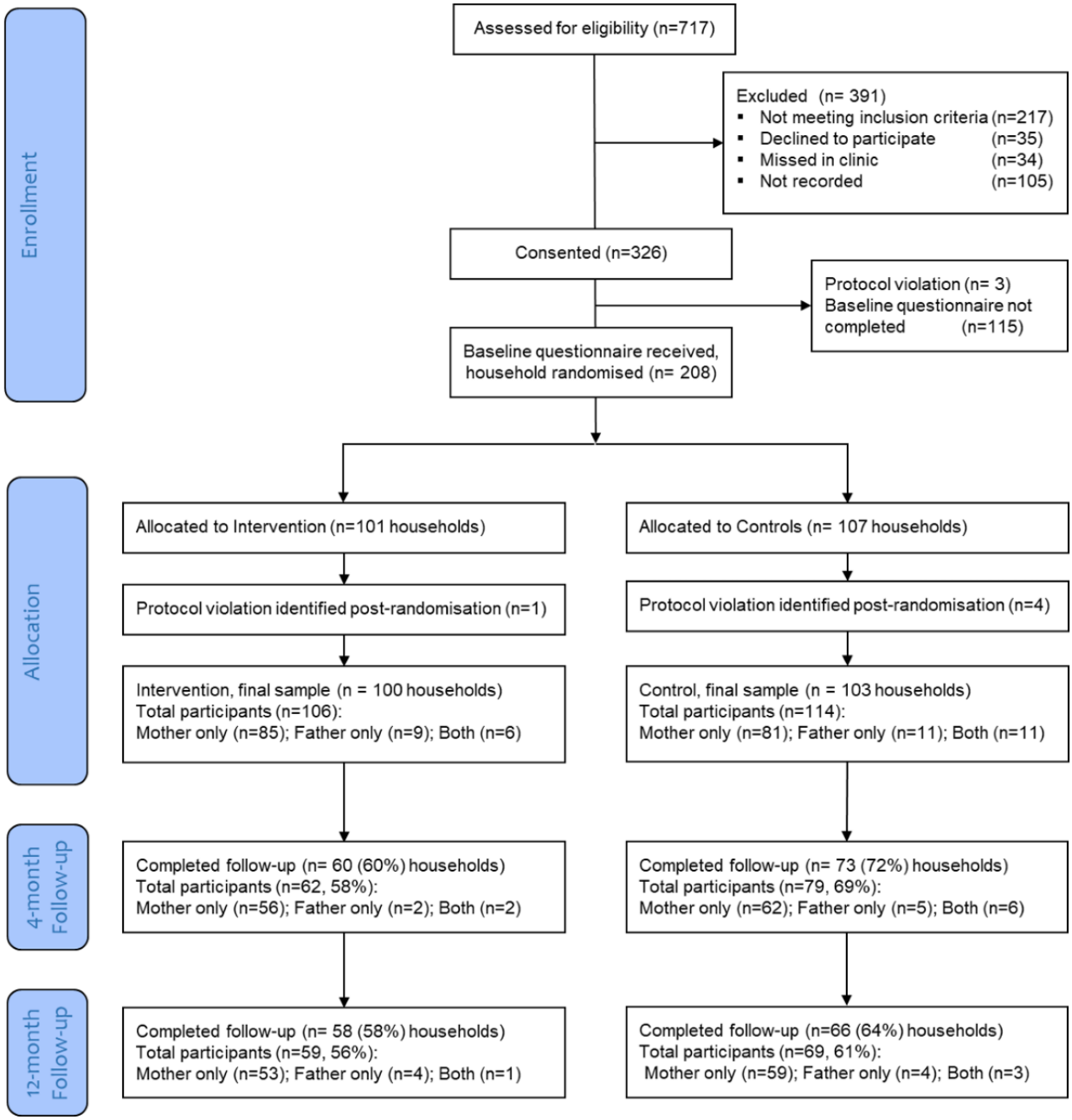
Overview

Between February 2016 and October 2018, 717 parents were assessed for eligibility and 326 (45.5%) consented to take part. Baseline questionnaires were returned by parents of 207 children

XSL•FO
RenderX

(*households*). In all, 4 protocol violations were identified (diagnosis not JIA, 2/4, 50%; >6 months since diagnosis, 1/4, 25%; and consent form not received, 1/4, 25%), leaving a final sample of 203 households (220 parents), 100 (49.3%) households (106/220, 48.2%, parents) randomized to the intervention arm and 103 (50.7%) households (114/220, 51.8%, parents) randomized to the control arm (Figure 1). Follow-up questionnaires were completed by 65.5% (133/203) of the households (141/220, 64.1%, parents) at 4 months and 61.1% (124/203) of the households (128/220, 58.2%, parents) at 12 months. Attrition did not differ significantly between intervention and control at 4 months (χ^2_1=2.8; *P*=.10) or 12 months (χ^2_1=0.5; *P*=.47). Responders (those who completed one or both follow-ups) differed from nonresponders (those who completed neither follow-up) on two baseline variables: mothers (146/183, 79.8%) responded proportionally more than fathers (20/37, 54.1%; B=1.210, SE 0.378; *P*=.001; odds ratio 3.354, 95% CI 1.600-7.033) and responders scored higher on the baseline ECS17-A subscale *use of health information* (mean 77.4, SE 1.26) than nonresponders (mean 71.0, SE 2.77; B=0.020, SE 0.009; *P*=.02; odds ratio 1.021, 95% CI 1.003-1.039).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Missing Values

At baseline, 58.8% (50/85) of the variables and 88.2% (194/220) of the cases had complete data, relating to an overall missing data level of 5.7% (Little MCAR test: χ^2_{7726}=7356.6; *P*=.99). At the 4-month follow-up, 83.7% (118/141) of the cases followed up had complete data but no variables were complete, relating to an overall missing data level of 9.4% (Little MCAR test: χ^2_{3333}=454.2; *P*=.99). At the 12-month follow-up, 82.8% (106/128) of the cases had complete data, with 87% (40/46) of

XSL•FO
RenderX

the variables having complete data, relating to an overall missing data level of 12.4% (Little MCAR test: χ^2_{2978}=140.3; *P*=.99).

Baseline Characteristics

Baseline characteristics are shown in Table 2. In 8.4% (17/203) of the cases, both parents took part. Most of the participants were mothers (183/220, 83.2%). The parents' average age was 36.5 (SD 6.5) years, and 31.8% (70/220) were educated to degree level or above. Participants' children with JIA were predominantly girls (136/203, 67%), with a mean age of 6.1 (SD 3.4) years. The most frequent JIA subtypes were oligoarticular (107/203, 52.7%) and polyarticular (65/203, 32%).

Unadjusted means and SEs for all questionnaires at each time point are presented in Multimedia Appendix 1. No difference between the trial arms was found on any clinical or self-report questionnaire at baseline; however, education level was higher in the intervention arm (mean 3.65, SD 1.63; control: mean 3.17, SD 1.52; $F_{1,218}$=5.221; *P*=.02). This was controlled for in all analyses.

XSL•FO
RenderX

Table 2. Participant characteristics at baseline.

	Intervention	Control	Total
Parents, details			
Households, n	100	103	203
Participants, n	106	114	220
Parents, n (%)			
Mother	91 (85.8)	92 (80.7)	183 (83.2)
Father	15 (14.2)	22 (19.3)	37 (16.8)
Age (years), mean (SD)	35.8 (6.5)	37.2 (6.4)	36.5 (6.5)
Education, n (%)			
≤GCSE[a] or equivalent	33 (31.2)	52 (45.6)	85 (38.7)
Advanced Level[b] or equivalent	26 (24.5)	25 (21.9)	51 (23.2)
HNC[c] or HND[d]	7 (6.6)	7 (6.1)	14 (6.4)
Degree or postgraduate	40 (37.7)	30 (26.3)	70 (31.8)
Relationship status, n (%)			
Single or divorced or separated	11 (10.4)	19 (16.6)	30 (13.6)
Married or living with partner or in a relationship	95 (89.6)	95 (83.3)	190 (86.4)
Living with child with JIA[e], n (%)	104 (98.1)	111 (97.4)	215 (97.7)
Average number of children per family, mean (SE)	2.11 (0.09)	2.03 (0.09)	2.07 (0.06)
Child with JIA, details			
Total, n	100	103	203
Gender, n (%)			
Female	69 (69)	67 (65)	136 (67)
Male	31 (31)	36 (35)	67 (33)
Age (years), mean (SD)	6.3 (3.2)	6.0 (3.7)	6.1 (3.4)
JIA subtype, n (%)			
Systemic	4 (4)	4 (3.9)	8 (3.9)
Oligoarticular	58 (58)	49 (47.6)	107 (52.7)
Polyarticular	28 (28)	37 (35.9)	65 (32)
Psoriatic	6 (6)	5 (4.9)	11 (5.4)
ERA[f]	3 (3)	4 (3.9)	7 (3.4)
Undifferentiated	1 (1)	4 (3.9)	5 (2.5)
Current disease severity, median (IQR)			
Number of active joints (known for 189/203, 93.1%)	2 (1-5)	2 (1-5.5)	2 (1-5)
Number of limited joints (known for 189/203, 93.1%)	2 (1-4)	2 (1-4)	2 (1-4)
CHAQ[g] (known for 128/203, 63.1%)	0.8 (0.3-1.3)	0.8 (0-1.4)	0.8 (0.1-1.3)
Parent global (known for 117/203, 57.6%)	3.3 (0.7-5.4)	2.8 (1.0-6.0)	3.0 (0.9-5.7)
Physician global (known for 113/203, 55.7%)	3.0 (1.5-6.0)	2.0 (0.9-5.0)	2.5 (1.0-5.0)
ESR[h] (known for 135/203, 66.5%)	17.0 (6.0-38.0)	22.5 (7.3-36.5)	20.2 (7.0-37.0)
Medication, n (%)			
Methotrexate	31 (31)	39 (37.9)	70 (34.5)
Biologic	0 (0)	2 (1.9)	2 (1)

XSL•FO
RenderX

[a]GCSE: General Certificate of Secondary Education, national exam taken at approximately age 16 years.

[b]Advanced Level: national exam taken at approximately age 18 years.

[c]HNC: higher national certificate.

[d]HND: higher national diploma.

[e]JIA: juvenile idiopathic arthritis.

[f]ERA: enthesitis-related arthritis.

[g]CHAQ: Child Health Assessment Questionnaire.

[h]ESR: erythrocyte sedimentation rate.

Baseline Questionnaire Data

The PIP asks about child illness–related events; therefore, there were no normative data from parents of healthy children. Baseline PIP-F (mean 108.45, SE 2.14; mothers: mean 111.04, SE 2.31; fathers: mean 95.68, SE 5.22) and PIP-D scores (mean 102.84, SE 2.10; mothers: mean 104.73, SE 2.28; fathers: mean 93.48, SE 5.06) were worse than those reported by a sample of UK and US parents of children with a history of heart disease [31] (PIP-F mean 80.3 for mothers; mean difference 30.736, SE 2.306; $t_{7353301}$=13.13; P<.001; and mean 76.2 for fathers; mean difference 19.483, SE 5.219; $t_{5175615}$=3.73; P<.001); PIP-D mean 80.6 for mothers; mean difference 24.132, SE 2.284; $t_{1063948}$=10.57; P<.001); and mean 75.7 for fathers; mean difference 17.776, SE 5.058; $t_{1806544}$=3.51; P<.001) and UK parents of adolescents with chronic pain [32] (PIP-F mean 104.9; mean difference 3.554, SE 2.140; $t_{6166319}$=1.66; P=.01) and PIP-D mean 98.0; mean difference 4.839, SE 2.097; t_{569347}=2.31; P=.02).

Baseline scores for Hospital Anxiety and Depression Scale anxiety and depression (mean 9.04, SD 0.34, and mean 5.49, SD 0.29, respectively) were significantly worse than published UK normative data [33] (anxiety: mean 6.14, SD 3.76; depression: mean 3.68, SD 3.07; anxiety: mean difference 2.896, SE 0.341; t_{62011}=8.49; P<.001; depression: mean difference 1.806, SE=0.285; t_{196469}=6.33; P<.001). Of the 220 participants, scoring in the *mild* (score 8-10), *moderate* (11-14), or *severe* (15-21) ranges for anxiety were 48 (21.8%), 47 (21.4%), and 34 (15.5%) participants, respectively, and for depression were 43 (19.5%), 24 (10.9%), and 3 (1.4%) participants, respectively. This compares with 19% of the women and 12.5% of the men scoring in the moderate to severe ranges for anxiety and 6.9% of both men and women scoring in the moderate to severe ranges for depression in a UK normative sample [34].

Baseline PASE self-efficacy scores were mean 4.44 (SE 0.15) for symptoms (mothers: mean 4.42, SE 0.16; fathers: mean 4.54, SE 0.39) and mean 5.80 (SE 0.15) for psychosocial (mothers: mean 5.78, SE 0.16; fathers: mean 5.92, SE 0.14). These are approximately at the scale midpoint and are worse than those reported in the original scale validation [22] by mothers (symptoms: mean difference –0.428, SE 0.157; t_{8485}=–2.73; P=.006; psychosocial: mean difference –0.620, SE 0.164; t_{3078}=–3.77; P<.001) but not by fathers (symptoms: mean difference 0.681, SE 0.385; t_{12222}=1.77; P=.08; psychosocial: mean difference –0.308, SE 0.414; t_{6936}=–0.74; P=.46).

The total score on the ECS17-A was mean 77.87 (SE 0.97). This score reflects that, on average, parents felt that they could *usually* manage their child's health care. The score on the Client Satisfaction Questionnaire of mean 28.4 (SE 0.24) reflects very high satisfaction with health services.

The CHQ-PF50 health-related quality of life summary scores of participants' children were mean 33.4 (SE 0.95) for physical quality of life and mean 44.4 (SE 0.72) for psychosocial quality of life, which are poorer than the scores reported for UK healthy controls [25] (mean 55.4, SD 4.2; mean difference –21.999, SE 0.947; t_{34421}=–23.23; P<.001 for physical quality of life; mean 51.6, SD 7.1; mean difference –7.199, SE 0.721; t_{27490}=–9.98; P<.001 for psychosocial quality of life).

Trial Outcomes

Overview

Tables 3 and 4 present adjusted means at each follow-up per group and multilevel modeling analyses estimates for the effect of trial arm and time and their interaction on all outcomes, adjusted for baseline scores and education level, respectively. The use of random effects for parent clusters was not possible because the number of dyad clusters was too few and random effects analyses did not converge. Consequently, parents were treated as individual units. Post hoc comparisons of trial arm effects at 4 months and 12 months are reported in Multimedia Appendix 2. The direction of effects, shown in Figures 2 and 3, mostly favored the intervention arm. Results for individual outcomes are reported in the next sections.

XSL•FO
RenderX

Table 3. Follow-up adjusted means (adjusted for baseline scores and educational level) on each outcome for the control and intervention groups in multilevel modeling analysis.

Variable	Control		Intervention	
	4 months, adjusted mean[a] (95% CI)	12 months, adjusted mean (95% CI)	4 months, adjusted mean (95% CI)	12 months, adjusted mean (95% CI)
PIP[b]frequency				
Communication	19.01 (17.86-20.16)	19.47 (18.27-20.68)	17.45 (16.16-18.74)	17.36 (16.05-18.67)
Medical care	18.29 (16.95-19.63)	18.48 (17.04-19.91)	16.80 (15.30-18.31)	16.63 (15.07-18.20)
Emotional distress	37.86 (35.65-40.07)	36.95 (34.68-39.21)	35.43 (32.95-37.90)	35.08 (32.60-37.57)
Role function	19.50 (18.24-20.76)	20.10 (18.77-21.42)	18.16 (16.75-19.57)	17.74 (16.28-19.21)
Frequency total	94.38 (89.04-99.73)	94.62 (89.13-100.11)	88.32 (82.32-94.31)	87.26 (81.27-93.24)
PIP difficulty				
Communication	17.64 (16.45-18.82)	18.13 (16.87-19.39)	15.82 (14.52-17.12)	15.57 (14.24-16.90)
Medical care	16.88 (15.53-18.23)	16.57 (15.21-17.93)	14.89 (13.37-16.41)	14.49 (12.98-16.00)
Emotional distress	38.87 (36.33-41.41)	38.47 (35.93-41.00)	35.85 (32.98-38.72)	35.08 (32.32-37.84)
Role function	19.15 (17.73-20.57)	19.39 (17.97-20.82)	17.57 (15.99-19.16)	17.10 (15.51-18.68)
Difficulty total	92.63 (86.86-98.40)	92.35 (86.77-97.93)	84.09 (77.60-90.58)	81.82 (75.76-87.87)
HADS[c]				
Anxiety	8.33 (7.57-9.09)	7.86 (7.01-8.71)	7.61 (6.75-8.46)	7.95 (7.02-8.87)
Depression	5.52 (4.82-6.21)	5.64 (4.93-6.35)	4.78 (4.00-5.57)	5.05 (4.27-5.84)
PASE[d]				
Symptoms	5.31 (4.92-5.69)	5.09 (4.65-5.52)	5.03 (4.57-5.49)	5.71 (5.24-6.18)
Psychosocial	6.58 (6.12-7.03)	6.31 (5.83-6.79)	6.55 (6.04-7.06)	6.64 (6.11-7.16)
ECS17-A[e]				
Use health information	76.70 (73.47-79.93)	78.32 (75.31-81.32)	75.07 (73.22-76.92)	81.04 (79.36-82.73)
Clarify priorities	82.89 (80.07-85.72)	83.65 (80.64-86.67)	81.94 (78.79-85.08)	85.62 (82.37-88.87)
Communicate with others	85.21 (82.29-88.14)	84.20 (81.32-87.08)	79.87 (78.19-81.55)	87.34 (85.74-88.94)
Negotiate roles	75.45 (72.12-78.79)	76.03 (72.63-79.42)	74.59 (70.85-78.32)	79.79 (76.10-83.49)
Decide and act	77.01 (73.82-80.20)	76.90 (73.97-79.83)	75.12 (71.53-78.70)	80.84 (77.64-84.04)
ESC17-A total	79.15 (76.66-81.64)	79.45 (76.90-82.01)	76.67 (73.88-79.46)	82.51 (79.73-85.29)
CSQ[f]	28.30 (27.51-29.09)	28.07 (27.30-28.84)	28.70 (27.81-29.59)	28.51 (27.67-29.35)
CHQ-PF50[g]				
Physical functioning	74.49 (69.83-79.15)	76.80 (71.94-81.66)	80.35 (75.10-85.61)	83.02 (77.73-88.30)
Role/social limitations–emotional/behavioral	80.79 (75.72-85.85)	83.62 (78.83-88.40)	87.94 (82.25-93.63)	87.22 (82.02-92.43)
Role/social limitations–physical	78.77 (73.63-83.92)	80.17 (75.03-85.31)	85.86 (80.06-91.66)	86.75 (81.16-92.35)
Bodily pain and discomfort	58.85 (54.06-63.63)	62.40 (57.02-67.77)	66.55 (61.16-71.94)	67.27 (61.40-73.14)
Behavior	67.15 (63.86-70.43)	67.19 (63.87-70.51)	66.61 (62.90-70.32)	69.66 (66.05-73.27)
Mental health	71.22 (68.40-74.04)	72.77 (69.46-76.08)	75.19 (72.03-78.34)	71.60 (68.01-75.20)
Self-esteem	70.06 (66.43-73.70)	76.42 (72.62-80.22)	76.23 (72.13-80.34)	77.02 (72.89-81.14)
General health perceptions	52.41 (49.31-55.52)	54.74 (51.39-58.09)	56.27 (52.78-59.77)	57.31 (53.65-60.97)
Parental impact–emotional	63.69 (58.95-68.44)	63.60 (58.99-68.20)	67.74 (62.38-73.10)	68.52 (63.50-73.54)
Parental impact–time	75.58 (70.84-80.32)	81.99 (77.65-86.34)	82.84 (77.49-88.19)	84.80 (80.10-89.51)
Family activities	71.69 (67.88-75.51)	73.47 (69.29-77.65)	77.01 (72.72-81.30)	79.09 (74.52-83.65)

XSL•FO
RenderX

Variable	Control		Intervention	
	4 months, adjusted mean[a] (95% CI)	12 months, adjusted mean (95% CI)	4 months, adjusted mean (95% CI)	12 months, adjusted mean (95% CI)
Family cohesion	76.27 (72.09-80.44)	74.03 (69.89-78.17)	78.53 (73.81-83.24)	79.89 (75.41-84.37)
Physical summary scores	39.62 (37.14-42.09)	41.56 (38.90-44.22)	43.06 (40.32-45.80)	44.51 (41.63-47.40)
Psychosocial summary scores	45.78 (44.00-47.56)	48.06 (46.15-49.96)	48.79 (46.78-50.79)	48.68 (46.61-50.75)

[a]Adjusted mean for baseline scores and educational level.

[b]PIP: Pediatric Inventory for Parents.

[c]HADS: Hospital Anxiety and Depression Scale.

[d]PASE: Parent's Arthritis Self-Efficacy Scale.

[e]ECS17-A: Effective Consumer Scale–Adapted.

[f]CSQ: Client Satisfaction Questionnaire.

[g]CHQ-PF50: Child Health Questionnaire, 50-item parent version.

XSL•FO
RenderX

Table 4. Multilevel modeling analyses of each outcome exploring time and trial arm main effects and their interactions.

Variable	Effect					
	Trial arm		Time		Time×trial arm	
	F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
PIP[a] frequency						
Communication	5.37 (1,120627)	.02	0.01 (1,225948)	.90	0.32 (1,218596)	.58
Medical care	2.87 (1,36383)	.09	0.04 (1,145323)	.84	0.01 (1,45835)	.75
Emotional distress	1.17 (1,69328)	.28	0.07 (1,115020)	.79	0.10 (1,129609)	.75
Role function	5.40 (1,27203)	.02	0.34 (1,89283)	.56	1.10 (1,67018)	.30
Frequency total	3.14 (1,7872787)	.08	0.12 (1,2123768)	.73	0.09 (1,3941957)	.76
PIP difficulty						
Communication	7.43 (1,2237)	.006	0.09 (1,6133)	.76	0.45 (1,5150)	.50
Medical care	4.04 (1,2907)	.04	0.20 (1,1715)	.66	0.01 (1,6475)	.94
Emotional distress	3.10 (1,6128)	.08	0.22 (1,7769)	.64	0.03 (1,15313)	.87
Role function	4.37 (1,821)	.04	0.25 (1,1028)	.62	0.31 (1,1590)	.58
Difficulty total	6.30 (1,588193)	.01	0.37 (1,115056)	.54	0.16 (1,194822)	.69
HADS[b]						
Anxiety	0.02 (1,3123)	.89	0.56 (1,6920)	.45	1.72 (1,5892)	.19
Depression	1.16 (1,11080)	.28	0.41 (1,4597)	.52	0.07 (1,12694)	.80
PASE[c]						
Symptoms	3.63 (1,10281)	.06	4.80 (1,961)	.03	4.90 (1,7284)	.03
Psychosocial	0.84 (1,7031)	.36	0.07 (1,12798)	.79	0.59 (1,3629)	.44
ECS17-A[d]						
Use health information	1.42 (1,4.68E+09)	.23	9.90 (1,7.90E+10)	.002	2.89 (1,4.56E+08)	.09
Clarify priorities	0.75 (1,1.95E+08)	.39	3.56 (1,8.41E+12)	.06	1.22 (1,3.55E+07)	.27
Communicate with others	2.08 (1,2.03E+09)	.15	14.66 (1,2.07E+11)	<.001	10.33 (1,2.07E+08)	.001
Negotiate roles	2.16 (1,1.40E+08)	.14	6.21 (1,1.33E+13)	.01	2.68 (1,5.77E+06)	.10
Decide and act	3.17 (1,7.55E+06)	.08	8.63 (1,1.79E+08)	.003	4.91 (1,1.22E+07)	.03
ECS17-A total	2.51 (1,2.29E+07)	.11	14.04 (1,7.75E+07)	<.001	6.90 (1,4.36E+06)	.009
CSQ[e]	0.57 (1,16736)	.45	0.16 (1,30483)	.69	0.00 (1,14841)	.95
CHQ-PF50[f]						
Physical functioning	2.88 (1,546164)	.09	0.77 (1,403462)	.38	0.01 (1,533107)	.93
Role/social limitations–emotional/behavioral	1.00 (1,360571)	.32	0.05 (1,454651)	.82	0.69 (1,577910)	.41
Role/social limitations–physical	2.89 (1,273546)	.09	0.07 (1,359145)	.79	0.01 (1,66209)	.91
Bodily pain and discomfort	1.44 (1,1039728)	.23	0.06 (1,579724)	.80	0.50 (1,777799)	.48
Behavior	0.97 (1,37269)	.33	1.98 (1,48284)	.16	1.05 (1,28885)	.31
Mental health	0.22 (1,96801)	.64	2.83 (1,264228)	.09	3.15 (1,83261)	.08
Self-esteem	0.04 (1,22019)	.84	0.10 (1,92848)	.75	2.73 (1,71775)	.10
General health perceptions	1.03 (1,198499)	.31	0.28 (1,353837)	.59	0.24 (1,174477)	.62
Parental impact–emotional	2.00 (1,199245)	.16	0.07 (1,394803)	.79	0.05 (1,434994)	.82
Parental impact–time	0.74 (1,650668)	.39	0.46 (1,434479)	.50	1.27 (1,237236)	.26

XSL•FO
RenderX

Variable	Effect					
	Trial arm		Time		Time×trial arm	
	F test (*df*)	*P* value	*F* test (*df*)	*P* value	*F* test (*df*)	*P* value
Family activities	3.13 (1,186376)	.08	0.77 (1,396224)	.38	0.01 (1,493035)	.93
Family cohesion	3.53 (1,225860)	.06	0.21 (1,222878)	.65	0.81 (1,321685)	.37
Physical summary scores	2.13 (1,8378)	.14	0.93 (1,27479)	.33	0.06 (1,8741)	.81
Psychosocial summary scores	0.19 (1,6310)	.66	0.01 (1,5005)	.93	2.31 (1,1647)	.13

[a]PIP: Pediatric Inventory for Parents.

[b]HADS: Hospital Anxiety and Depression Scale.

[c]PASE: Parent's Arthritis Self-Efficacy Scale.

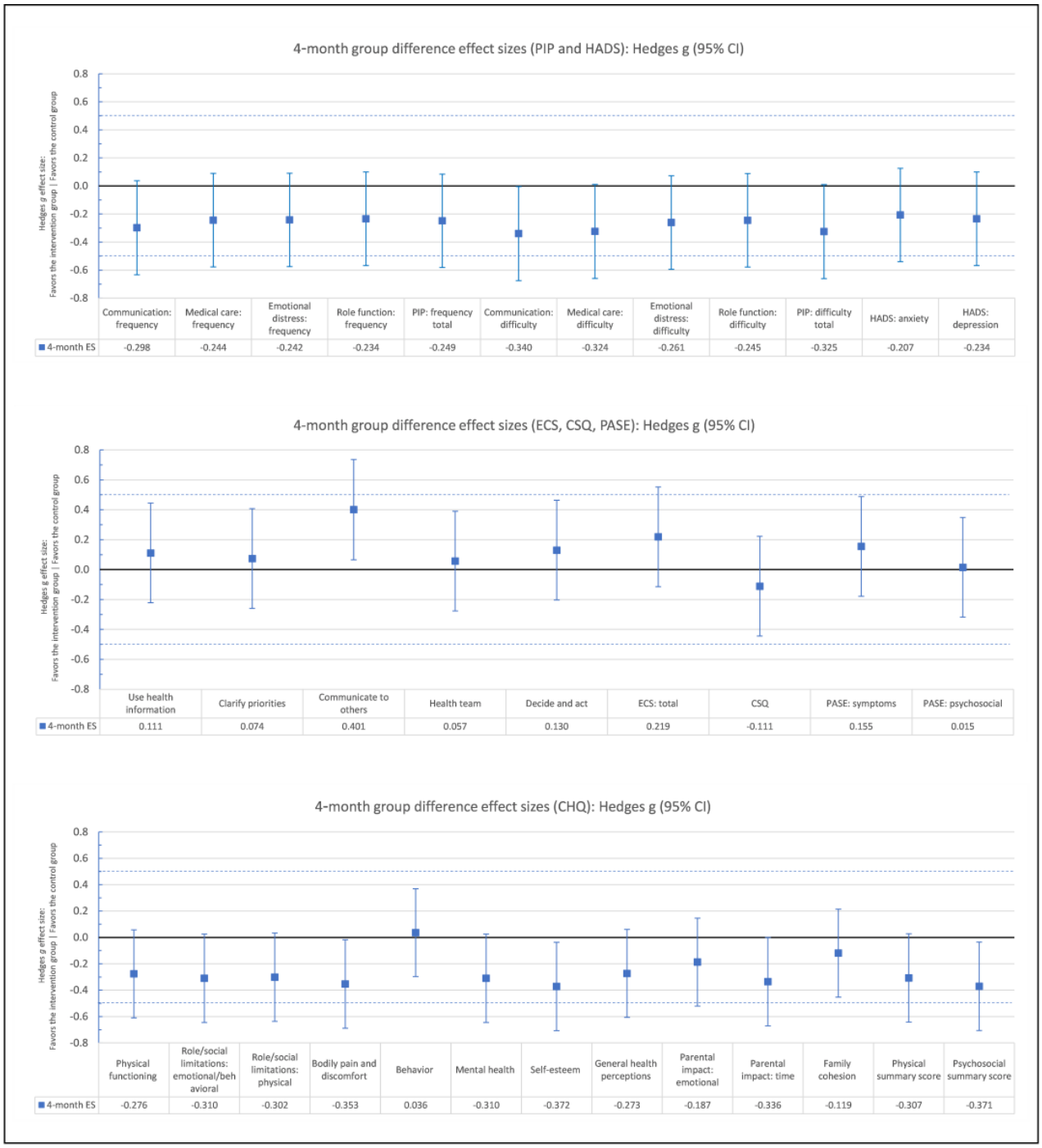
[d]ECS17-A: Effective Consumer Scale–Adapted.

[e]CSQ: Client Satisfaction Questionnaire.

[f]CHQ-PF50: Child Health Questionnaire, 50-item parent version.

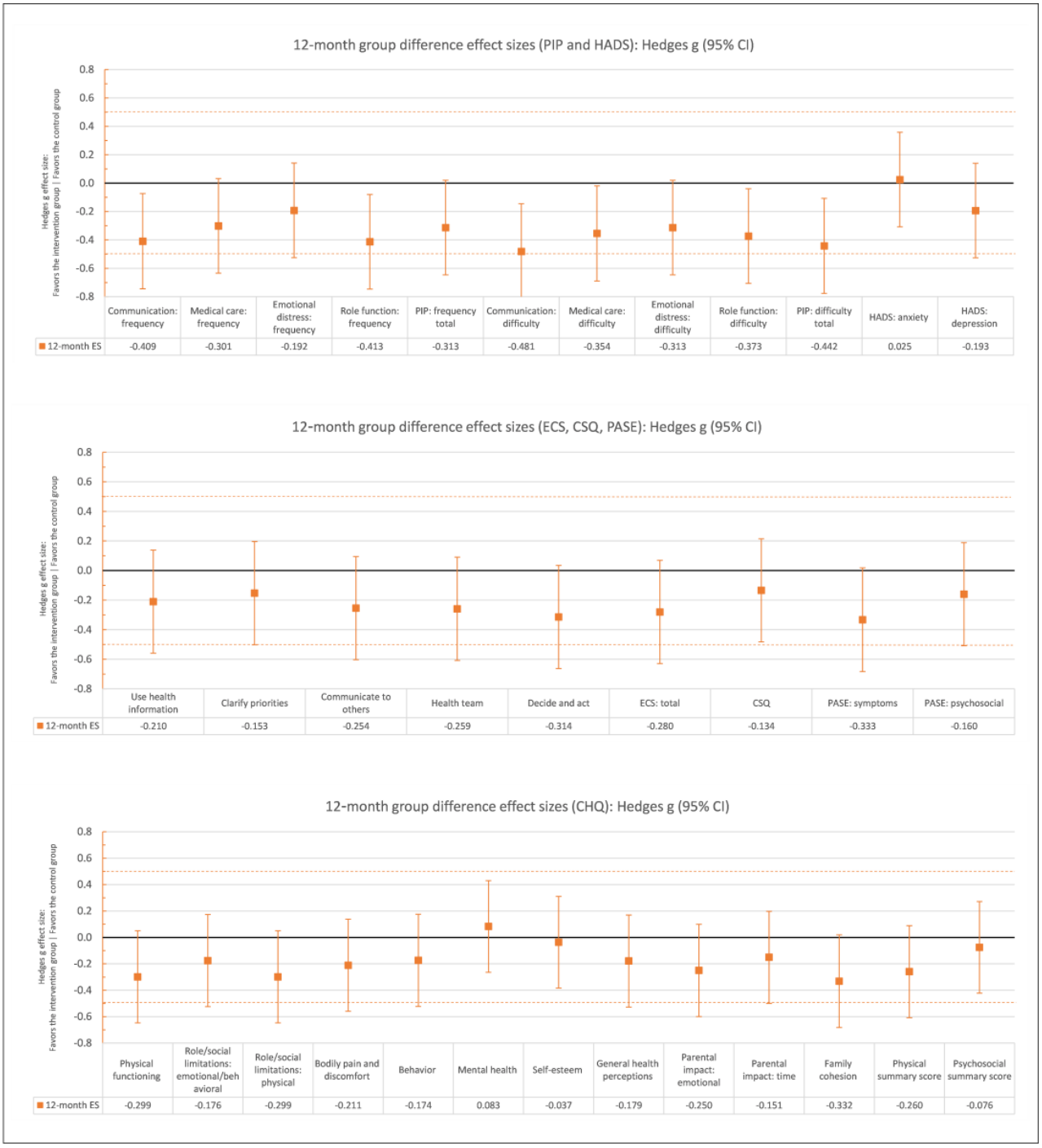
XSL•FO
RenderX

Figure 2. Group difference effect sizes at 4 months after randomization for all trial outcomes.CHQ: Child Health Questionnaire; CSQ: Client Satisfaction Questionnaire; ECS: Effective Consumer Scale; ES: effect size; HADS: Hospital Anxiety and Depression Scale; PASE: Parent's Arthritis Self-Efficacy Scale; PIP: Pediatric Inventory for Parents.



XSL•FO
RenderX

Figure 3. Group difference effect sizes at 12 months after randomization for all trial outcomes.CHQ: Child Health Questionnaire; CSQ: Client Satisfaction Questionnaire; ECS: Effective Consumer Scale; ES: effect size; HADS: Hospital Anxiety and Depression Scale; PASE: Parent's Arthritis Self-Efficacy Scale; PIP: Pediatric Inventory for Parents



Parenting Stress

A significant effect of the trial arm over the 2 follow-up periods was found on the PIP-F subscales *communication* and *role function* and on the PIP-D subscales *communication*, *medical care*, and *role function*, as well as the PIP-D total score. In each instance, participants in the intervention arm reported less frequency and difficulty of illness-related stressful events than participants in the control arm. Post hoc comparisons (Multimedia Appendix 2) found that these effects mostly reached statistical significance at 12 months. Effect sizes were small to medium.

Anxiety and Depression

We did not find a significant effect of the intervention on mean anxiety or depression scores.

Arthritis Self-efficacy

No significant effect of the trial arm was found on PASE. However, there was a significant time effect on PASE symptoms, where the whole sample improved over the 12-month period. In addition, there was a significant interaction effect on PASE symptoms: participants in the intervention arm reported

XSL•FO
RenderX

greater improvement in their self-efficacy from 4 to 12 months than control participants.

Perceived Effectiveness in Managing Health Care

We did not find an overall effect of the trial arm on parents' perceived effectiveness in managing their child's health care assessed with the ECS17-A. We found significant interaction effects on the subscales *communicating with others* and *deciding and taking action*. For *communicating with others*, there was a significant trial arm effect at 4 months favoring the control arm, but the control arm did not change between 4 and 12 months, whereas the intervention arm improved significantly. For *deciding and taking action*, there was no effect of the trial arm at 4 or 12 months and the control arm did not change between 4 and 12 months, but the intervention arm improved significantly. We found a main effect of time on the subscales *use of health information* and *negotiating roles and taking control*, with an improvement in the overall sample on both subscales. There was no effect on the subscale *clarifying personal priorities*.

Satisfaction With Health Care

Satisfaction with health care services was very high at baseline and remained so at follow-ups with no significant differences between the trial arms.

Child's Health-Related Quality of Life

There was no significant overall effect of the trial arm on parents' assessment of their child's health-related quality of life on the CHQ-PF50.

Discussion

Principal Findings

This RCT evaluated the *WebParC* website for parents of children with JIA. To our knowledge, this is the first website for parents of children with JIA that has undergone evaluation in an RCT. The website was found to be successful in reducing child illness–related parenting stress and also promoted a greater improvement in parents' self-efficacy in managing children's symptoms.

Although the direction of effects mostly favored the intervention, post hoc comparisons indicated that they did not reach statistical significance until 12 months. This suggests that it is in the longer term that the knowledge and skills parents gain from the website significantly reduce their stress and improve symptom self-efficacy.

Satisfaction with health care was very high among parents throughout this trial, indicating that even in the context of excellent clinical care, parents experience stress related to their child's illness. This trial has shown that a web-based intervention, accessible when needed outside of the clinical setting, can help parents to manage the stress of having a child with JIA and could be offered to parents as an adjunct to the care given to their child. The effect sizes achieved ranged from small to medium, which is acceptable for a very *light touch* intervention that demands few additional resources.

Scores on the ECS17-A subscale *communicating with others* were high at all time points, reflecting a good degree of confidence in communicating with the health care team across the trial period. However, the intervention arm scores deteriorated at 4 months before improving again at 12 months. The drop at 4 months may indicate that access to the website meant that parents were less likely to engage with health care professionals in the early stage but had more interaction and had built up confidence in the longer term.

The parent outcomes improved by the website were those relating to the stress of communication; managing medical aspects of their child's care, including symptoms; and carrying out everyday family and social roles. It is important that the website, which covers information about JIA and its treatment, including potentially distressing issues such as medication side effects, did not have any negative effect on parents' psychological well-being. Of the 3 main *tasks* in living with a chronic illness proposed by Corbin and Strauss [35], two were improved by WebParC: medical management and role management. The third task, managing emotions, was unchanged.

We were unable to identify evaluations of other interventions specifically for parents of children with JIA. A review of interventions for parenting stress in families with pediatric conditions [35] did not include any web-based interventions. A Cochrane review of 47 psychological interventions for parents of children with chronic illnesses [37] included 6 interventions that were delivered at least partly on the web. Of these, only 2 small trials (n<40) assessed parental mental health; therapist-supported web-based family problem solving [38] for traumatic brain injury was found to be beneficial, but part–web-based cognitive behavioral therapy [39] did not have an effect on the mental health of parents who had a child with cancer. The primary target of WebParC was parenting stress rather than mental health; using web-based approaches to support the mental health of parents of children with JIA may require a greater focus on parents' psychological well-being than we were able to achieve in WebParC.

Limitations of the study, in common with interventions of this type, include that it was not possible to blind participants to trial arm allocation. Although requested not to inform their child's clinicians of their allocation, it is not possible to know whether all participants followed this request. We made every effort to ensure that where both parents participated, questionnaires were given to the individual parent for completion. Although we consider it unlikely, it is nonetheless possible that 1 parent completed both copies. However, the number of questionnaires received from both parents is small. A proportion of parents who consented to participate in the trial did not return the baseline questionnaire and were therefore not randomized. The follow-up response rates were also lower than expected. Parents may have forgotten or not prioritized questionnaire completion; when reminders and a small incentive were introduced midtrial, rates of baseline and follow-up questionnaire return improved. Another possibility is that although parents consented to the trial when they were at the clinic, taking part in research about their own well-being, rather than their child's, was not a priority for them. Parents who did

XSL•FO
RenderX

not return the baseline questionnaire may also have been reluctant to answer detailed questions about their own and their child's well-being. These issues will need to be considered in future studies of this type.

In common with other research [40], fathers were less likely to participate, which occurred in this trial because they were less likely than mothers to attend the clinic. It was not possible to establish whether nonparticipating fathers may have been given access to the website by participating partners. The small number of cases where both parents participated meant that we were unable to cluster by household in our analyses.

To minimize participant burden, outcomes were assessed at only two follow-up times, 4 months and 12 months, after randomization. These were chosen for pragmatic reasons and to allow parents time to use the website before assessing its impact in the short and medium-to-longer term. We acknowledge that we will not have been able to capture all potential stressors that may have occurred and coping strategies used between the baseline and follow-up periods, but more frequent assessments would have increased the burden on parents.

Primary analyses with the PIP scales used a *P* value of <.05 for significance as per the protocol. For secondary analyses on additional scales we did not adjust the *P* value of <.05 to allow for multiple testing; therefore, caution should be taken when interpreting the results. However, it is notable that the pattern of our findings, even where not statistically significant, were mostly in the direction favoring the Intervention arm; therefore, it is unlikely that our significant findings reflect type I error.

Conclusions

In conclusion, the study reported in this paper has demonstrated that web-based interventions for parents of children with JIA that combine information and skills training can result in significant benefits for parents. The benefits of reduced illness-related parenting stress and improvements in confidence and self-efficacy regarding parenting skills are important for a group not often seen as a high priority in health care. Future studies should attempt to devise techniques that reduce the loss to follow-up that was higher than projected in this study. In general, web-based interventions for parents of children with a chronic illness should be made a priority because they are easy to access at any time, replicable, and can offer a preventive approach to a large number of parents.

Acknowledgments

This paper presents independent research funded by the National Institute for Health Research (NIHR) under its Research for Patient Benefit Program (grant reference number PB-PG-1013-32005) and the University College London Hospitals Charity, Shipley-Rudge Fund. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. LRW is additionally supported by grants from the NIHR (to the NIHR Biomedical Research Centre at Great Ormond Street Hospital), Versus Arthritis (20164, 21593), and the UK Research and Innovation Medical Research Council (MR/R013926/1). The authors would like to thank the parents and health care professionals who helped to develop the website content. The authors would also like to thank all the parents who took part in the trial, the study sites, teams, and colleagues who recruited participants and collected clinical data, as well as the trial steering committee who provided invaluable advice. The health care professionals from Great Ormond Street Hospital and University College London Hospital involved in developing the website content were Muthana Al Obaidi, Katie Austin, Emily Boulter, Clive Edelsten, Despina Eleftheriou, Jill Ferrari, Sarah Francis, Yvonne Glackin, Ellie Haggart, Gillian Hardman, Sally Harris, Abdul Hassan, Rengasamy Janarthanan, Alison Kelly, Sandrine Lacassagne, Netali Levi, Polly Livermore, Sue Maillard, Neil Martin, Ruth McGowan, Elena Moraitis, Kiran Nistala, Charris Papadopoulou, Reshma Pattani, Clarissa Pilkington, Chris Rawcliff, Debajit Sen, Lucy R Wedderburn, and Karen Wynne.

The members of the WebParC Investigator Group are Dr Eslam Al-Abadi, Birmingham Children's Hospital; Dr Muthana Al-Obaidi, Department of Paediatric Rheumatology, Great Ormond Street Hospital NHS Foundation Trust, Infection, Inflammation and Rheumatology Section; NIHR Biomedical Research Centre, UCL Great Ormond Street Institute of Child Health; Dr Kate Armon, Addenbrooke's Hospital, Cambridge and Norfolk & Norwich University Hospital; Dr Kathy Bailey, Oxford University Hospitals; Lyndsey Bibb, New Cross Hospital, Wolverhampton; Dr Richard Brough, Shrewsbury & Telford Hospital NHS Trust; Katrin Buerkle, City, University of London; Jo Bytham, Addenbrooke's Hospital, Cambridge; Asyah Chhibda, Leeds Teaching Hospitals NHS Trust; Dr Alice Chieng, Royal Manchester Children's Hospital; Heather Chisem, Sheffield Children's NHS Foundation Trust; Louise Coke, Norfolk & Norwich University Hospital; Dr Karen Davies, The Robert Jones and Agnes Hunt Orthopaedic Hospital; Dr Hans de Graaf, Southampton Children's Hospital; Rebecca Denyer, New Cross Hospital, Wolverhampton; Kirsty Devine, Great North Children's Hospital, Newcastle; Annette Duggan, Royal Manchester Children's Hospital; Andrea Edwards, Addenbrooke's Hospital, Cambridge; Louisa Fear, Norfolk & Norwich University Hospital; Ruth Finch, Oxford University Hospitals; Dr Chris Flood, London South Bank University; Elizabeth Fofana, Southampton Children's Hospital; Sara Foster, Addenbrooke's Hospital, Cambridge; Sarah Hanson, Leeds Teaching Hospitals NHS Trust; Dr Sally Harris, Brighton and Sussex University Hospitals NHS Trust; Dr Kathryn Harrison, Birmingham Children's Hospital and University Hospital Coventry & Warwickshire; Dr Daniel P Hawley, Sheffield Children's NHS Foundation Trust; Eleanor Heaf, Royal Manchester Children's Hospital; Dr Shashivadan Hirani, City, University of London; Ruth Howman, Birmingham Children's Hospital; James Jones, Shrewsbury and Telford Hospital NHS Trust; Ruth Jones, Royal Stoke University Hospital; Dr Alice Leahy, Southampton Children's Hospital; Dr Valentina Leone, Leeds Teaching Hospitals NHS Trust; Gail Lindsay, Addenbrooke's Hospital, Cambridge; Carol Lydon, Alder Hey Children's Hospital; Ian MacDonald, Great Ormond Street Hospital; Emma MacLeod, Southampton

XSL•FO
RenderX

Children's Hospital; Dr Liza J McCann, Alder Hey Children's Hospital, Liverpool; Dr Flora McErlane, Great North Children's Hospital, Newcastle; Ann McGovern, Royal Manchester Children's Hospital; Diane Miller, Oxford University Hospitals; Dr Kathleen Mulligan, City, University of London; Maxine Mutten, Sheffield Children's NHS Foundation Trust; Professor Stanton P Newman, City, University of London; Tracy Oliver, Norfolk & Norwich University Hospital; Dr Jonathan Packham, Royal Stoke University Hospital; Helen Parker, Royal Stoke University Hospital; Sarrah Peerbux, City, University of London; Dr Clive Ryder, University Hospital Coventry & Warwickshire; Hilary Shepley, The Robert Jones and Agnes Hunt Orthopaedic Hospital; Dr Brian Shields, University Hospital Coventry & Warwickshire; Dr Taunton Southwood, Birmingham Children's Hospital; Rebecca Steele, University Hospital Coventry & Warwickshire; Dr Alexandra Tabor, Royal Stoke University Hospital; Jo Taylor, City, University of London; Lowri Thomas, Sheffield Children's NHS Foundation Trust; Fiona Thompson, University Hospital Coventry & Warwickshire; Louise Turner, Leeds Teaching Hospitals NHS Trust; Sarah Turner, The Robert Jones and Agnes Hunt Orthopaedic Hospital; Susan Wadeson, Alder Hey Children's Hospital; Diane Walia, Great North Children's Hospital, Newcastle; Joanna Watts, Norfolk & Norwich University Hospital; Nicola Watts, University Hospital Coventry & Warwickshire; Professor Lucy R Wedderburn, Great Ormond Street Hospital; Pauline Whitelaw, City, University of London.

Authors' Contributions

All authors were involved in drafting or revising this paper, and all approved the final version. SPH had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. SN, KM, SPH, LW, and SH were responsible for the study conception and design. JT, KM, and the WebParC Investigator Group were responsible for the acquisition of data. SPH, KM, and SN analyzed and interpreted the data.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Unadjusted means and SEs for all self-report questionnaire measures at baseline, 4 months, and 12 months.
[DOCX File , 57 KB - jmir_v24i5e29787_app1.docx]

Multimedia Appendix 2
Post hoc comparisons of trial arm effects, adjusted for baseline scores and educational level.
[DOCX File , 71 KB - jmir_v24i5e29787_app2.docx]

Multimedia Appendix 3
CONSORT-EHeatlh V1.6.1.
[PDF File (Adobe PDF File), 1200 KB - jmir_v24i5e29787_app3.pdf]

References

1. Mulligan K, Etheridge A, Kassoumeri L, Wedderburn LR, Newman S. Do mothers and fathers hold similar views about their child's arthritis? Arthritis Rheum 2009;61(12):1712-1718 [FREE Full text] [doi: 10.1002/art.25008] [Medline: 19950321]
2. Mulligan K, Kassoumeri L, Etheridge A, Moncrieffe H, Wedderburn LR, Newman S. Mothers' reports of the difficulties that their children experience in taking methotrexate for Juvenile Idiopathic Arthritis and how these impact on quality of life. Pediatr Rheumatol Online J 2013;11(1):23 [FREE Full text] [doi: 10.1186/1546-0096-11-23] [Medline: 23714105]
3. Buerkle KM, Hirani SP, Kassoumeri L, Etheridge A, Wedderburn LR, Newman SP. The impact of response to MTX on the illness and treatment beliefs of parents of children with JIA. Rheumatology 2011;50(Supplement 3):131.
4. Gómez-Ramírez O, Gibbon M, Berard R, Jurencak R, Green J, Tucker L, et al. A recurring rollercoaster ride: a qualitative study of the emotional experiences of parents of children with juvenile idiopathic arthritis. Pediatr Rheumatol Online J 2016;14(1):13 [FREE Full text] [doi: 10.1186/s12969-016-0073-9] [Medline: 26961237]
5. Cousino MK, Hazen RA. Parenting stress among caregivers of children with chronic illness: a systematic review. J Pediatr Psychol 2013;38(8):809-828. [doi: 10.1093/jpepsy/jst049] [Medline: 23843630]
6. Anthony KK, Bromberg MH, Gil KM, Schanberg LE. Parental perceptions of child vulnerability and parent stress as predictors of pain and adjustment in children with chronic arthritis. Child Hlth Care 2011;40(1):53-69. [doi: 10.1080/02739615.2011.537938]
7. Fair DC, Rodriguez M, Knight AM, Rubinstein TB. Depression and anxiety in patients with juvenile idiopathic arthritis: current insights and impact on quality of life, a systematic review. Open Access Rheumatol 2019;11:237-252 [FREE Full text] [doi: 10.2147/OARRR.S174408] [Medline: 31807093]
8. Kazak AE. Pediatric Psychosocial Preventative Health Model (PPPHM): research, practice, and collaboration in pediatric family systems medicine. Fam Syst Health 2006;24(4):381-395. [doi: 10.1037/1091-7527.24.4.381]

XSL•FO
RenderX

9. Wright J, Curran J, Rose-Davis B, Cellucci T, Duffy CM, Tucker LB, et al. Parental perspectives about research and knowledge translation in juvenile idiopathic arthritis. ACR Open Rheumatol 2020;2(3):138-146 [FREE Full text] [doi: 10.1002/acr2.11111] [Medline: 31997575]
10. Tonsaker T, Bartlett G, Trpkov C. Health information on the Internet: gold mine or minefield? Can Fam Physician 2014;60(5):407-408 [FREE Full text] [Medline: 24828994]
11. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are parents getting it right? A survey of parents' Internet use for children's health care information. Interact J Med Res 2015;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
12. Stinson JN, Tucker L, Huber A, Harris H, Lin C, Cohen L, et al. Surfing for juvenile idiopathic arthritis: perspectives on quality and content of information on the Internet. J Rheumatol 2009;36(8):1755-1762. [doi: 10.3899/jrheum.081010] [Medline: 19531749]
13. JIA website for parents. URL: http://www.webparc.org.uk/Account?ReturnUrl=%2f [accessed 2021-11-30]
14. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, International League of Associations for Rheumatology. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. J Rheumatol 2004;31(2):390-392. [Medline: 14760812]
15. Living with arthritis: caring for someone. Arthritis Care. URL: http://www.arthritiscare.org.uk/LivingwithArthritis/Caringforsomeone
16. Arthritis: childhood. Birmingham Arthritis Research Centre. URL: https://www.barc.org.uk/arthritis/childhood/index.html
17. Childrens Chronic Arthritis Association. URL: https://www.ccaa.org.uk/?id=2
18. Information on paediatric rheumatic diseases. Paediatric Rheumatology INternational Trials Organisation. URL: https://tinyurl.com/3zz4n77c
19. Juvenile idiopathic arthritis. Patient UK. URL: https://patient.info/bones-joints-muscles/rheumatoid-arthritis-leaflet/juvenile-idiopathic-arthritis
20. Streisand R, Braniecki S, Tercyak KP, Kazak AE. Childhood illness-related parenting stress: the pediatric inventory for parents. J Pediatr Psychol 2001;26(3):155-162. [doi: 10.1093/jpepsy/26.3.155] [Medline: 11259517]
21. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
22. Barlow JH, Shaw KL, Wright CC. Development and preliminary validation of a self-efficacy measure for use among parents of children with juvenile idiopathic arthritis. Arthritis Care Res 2000;13(4):227-236. [doi: 10.1002/1529-0131(200008)13:4<227::aid-anr7>3.0.co;2-n] [Medline: 14635277]
23. Kristjansson E, Tugwell PS, Wilson AJ, Brooks PM, Driedger SM, Gallois C, et al. Development of the effective musculoskeletal consumer scale. J Rheumatol 2007;34(6):1392-1400. [Medline: 17552066]
24. Larsen DL, Attkisson CC, 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]
25. Nugent J, Ruperto N, Grainger J, Machado C, Sawhney S, Baildam E, Paediatric Rheumatology International Trials Organisation. The British version of the Childhood Health Assessment Questionnaire (CHAQ) and the Child Health Questionnaire (CHQ). Clin Exp Rheumatol 2001;19(4 Suppl 23):S163-S167. [Medline: 11510323]
26. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997;4(1):92-100. [doi: 10.1207/s15327558ijbm0401_6] [Medline: 16250744]
27. Broadbent E, Petrie KJ, Main J, Weinman J. The brief illness perception questionnaire. J Psychosom Res 2006;60(6):631-637. [doi: 10.1016/j.jpsychores.2005.10.020] [Medline: 16731240]
28. Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, Martini A. Preliminary definition of improvement in juvenile arthritis. Arthritis Rheum 1997;40(7):1202-1209. [doi: 10.1002/1529-0131(199707)40:7<1202::AID-ART3>3.0.CO;2-R] [Medline: 9214419]
29. Little RJ, Rubin DB. Statistical analysis with missing data. New York, NY: John Wiley; 2002.
30. Rubin DB. Multiple imputation and non-response in surveys. New York, NY: John WIley; 1987.
31. Wray J, Cassedy A, Ernst MM, Franklin RC, Brown K, Marino BS. Psychosocial functioning of parents of children with heart disease-describing the landscape. Eur J Pediatr 2018;177(12):1811-1821. [doi: 10.1007/s00431-018-3250-7] [Medline: 30232593]
32. Cohen LL, Vowles KE, Eccleston C. Parenting an adolescent with chronic pain: an investigation of how a taxonomy of adolescent functioning relates to parent distress. J Pediatr Psychol 2010;35(7):748-757. [doi: 10.1093/jpepsy/jsp103] [Medline: 19914956]
33. Crawford JR, Henry JD, Crombie C, Taylor EP. Normative data for the HADS from a large non-clinical sample. Br J Clin Psychol 2001;40(4):429-434. [doi: 10.1348/014466501163904] [Medline: 11760618]
34. Breeman S, Cotton S, Fielding S, Jones GT. Normative data for the hospital anxiety and depression scale. Qual Life Res 2015;24(2):391-398. [doi: 10.1007/s11136-014-0763-z] [Medline: 25064207]
35. Corbin JM, Strauss AL. Unending work and care: managing chronic illness at home. San Francisco, CA: Jossey-Bass; 1988.

XSL•FO
RenderX

36. Golfenshtein N, Srulovici E, Deatrick JA. Interventions for reducing parenting stress in families with pediatric conditions: an integrative review. J Fam Nurs 2016;22(4):460-492. [doi: 10.1177/1074840716676083] [Medline: 27821622]
37. Eccleston C, Fisher E, Law E, Bartlett J, Palermo TM. Psychological interventions for parents of children and adolescents with chronic illness. Cochrane Database Syst Rev 2015;4(4):CD009660 [FREE Full text] [doi: 10.1002/14651858.CD009660.pub3] [Medline: 25874881]
38. Wade SL, Carey J, Wolfe CR. An online family intervention to reduce parental distress following pediatric brain injury. J Consult Clin Psychol 2006;74(3):445-454. [doi: 10.1037/0022-006X.74.3.445] [Medline: 16822102]
39. Marsland AL, Long KA, Howe C, Thompson AL, Tersak J, Ewing LJ. A pilot trial of a stress management intervention for primary caregivers of children newly diagnosed with cancer: preliminary evidence that perceived social support moderates the psychosocial benefit of intervention. J Pediatr Psychol 2013;38(4):449-461 [FREE Full text] [doi: 10.1093/jpepsy/jss173] [Medline: 23341547]
40. Macfadyen A, Swallow V, Santacroce S, Lambert H. Involving fathers in research. J Spec Pediatr Nurs 2011;16(3):216-219. [doi: 10.1111/j.1744-6155.2011.00287.x] [Medline: 21702882]

Abbreviations

CHQ-PF50: Child Health Questionnaire, 50-item parent version
ECS17-A: Effective Consumer Scale–Adapted
JIA: juvenile idiopathic arthritis
MCAR: missing completely at random
PASE: Parent's Arthritis Self-Efficacy Scale
PIP: Pediatric Inventory for Parents
PIP-D: Pediatric Inventory for Parents, difficulty subscale
PIP-F: Pediatric Inventory for Parents, frequency subscale
RCT: randomized controlled trial

Edited by R Kukafka; submitted 20.04.21; peer-reviewed by J Nunez-Cordoba, A Coda; comments to author 05.11.21; revised version received 10.12.21; accepted 17.12.21; published 12.05.22.

Please cite as:
Mulligan K, Hirani SP, Harris S, Taylor J, Wedderburn LR, Newman S, WebParC Investigator group
The Effects of a Web-Based Tool for Parents of Children With Juvenile Idiopathic Arthritis: Randomized Controlled Trial
J Med Internet Res 2022;24(5):e29787
URL: https://www.jmir.org/2022/5/e29787
doi:10.2196/29787
PMID:35551065

©Kathleen Mulligan, Shashivadan P Hirani, Sally Harris, Jo Taylor, Lucy R Wedderburn, Stanton Newman, WebParC Investigator group. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2022. 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.

XSL•FO
RenderX

Original Paper

A Gamified Personalized Normative Feedback App to Reduce Drinking Among Sexual Minority Women: Randomized Controlled Trial and Feasibility Study

Sarah C Boyle[1], PhD; Joseph W LaBrie[1], PhD; Bradley M Trager[1], PhD; Lauren D Costine[2], PhD

[1]HeadsUp Labs, Department of Psychology, Loyola Marymount University, Los Angeles, CA, United States

[2]Awaken Therapy Center, Los Angeles, CA, United States

Corresponding Author:
Sarah C Boyle, PhD
HeadsUp Labs
Department of Psychology
Loyola Marymount University
1 LMU Drive
Suite 3787
Los Angeles, CA, 90045
United States
Phone: 1 310 568 6681
Email: sarah.boyle@lmu.edu

Abstract

Background: Sexual minority women disproportionately engage in heavy drinking and shoulder the burden of alcohol dependence. Although several intensive interventions are being developed to meet the needs of treatment-seeking sexual minority women, there remains a lack of preventive interventions to reduce drinking and its consequences among women not yet motivated to reduce their alcohol consumption.

Objective: We aimed to examine the feasibility and efficacy of reducing alcohol-related risks via personalized normative feedback (PNF) on alcohol use and coping delivered within LezParlay, a social media–inspired digital competition designed to challenge negative stereotypes about lesbian, bisexual, and queer (LBQ)–identified sexual minority women.

Methods: Feasibility was assessed by examining engagement with LezParlay outside the context of an incentivized research study, assessing the characteristics of the LBQ women taking part, and examining the competition's ability to derive risk-reducing actual norms as well as levels of acceptability and perceived benefits reported by participants. Intervention efficacy was examined by randomizing a subsample of 499 LBQ alcohol consumers (ie, drinkers) already taking part in the competition to receive sexual identity–specific PNF on alcohol use and coping, alcohol use only, or control topics over only 2 rounds of play. Changes in alcohol use and negative consequences were examined 2 and 4 months after the delivery of treatment PNF.

Results: A total of 2667 diverse LBQ women played ≥1 round of LezParlay. The competition attracted large numbers of moderate and heavy drinkers; however, risk-reducing actual norms could still be derived from competition rounds and featured in PNF. Efficacy results revealed that drinkers who received PNF on alcohol use and both alcohol use and coping had similar reductions in their weekly drinks (P=.003; P<.001), peak drinks (P<.001; P<.001), and negative consequences (P<.001; P<.001) relative to those who received PNF on control topics at the 2-month follow-up. However, at the 4-month follow-up, reductions in alcohol consumption outcomes faded among those who received alcohol PNF only (weekly: P=.06; peak: P=.11), whereas they remained relatively robust among those who received PNF on both alcohol use and coping (weekly: P=.02; peak: P=.03). Finally, participants found the competition highly acceptable and psychologically beneficial as a whole.

Conclusions: The LezParlay competition was found to be a feasible and efficacious means of reducing alcohol-related risks in this population. Our findings demonstrate the utility of correcting sexual identity–specific drinking and coping norms to reduce alcohol-related risks among LBQ women and suggest that this approach may also prove fruitful in other stigmatized health disparity populations. To engage these populations in the real world and expand the psychological benefits associated with PNF, our findings also point to packaging PNF within a broader, culturally tailored competition designed to challenge negative group stereotypes.

Trial Registration: ClinicalTrials.gov NCT03884478; https://clinicaltrials.gov/ct2/show/NCT03884478

XSL•FO
RenderX

International Registered Report Identifier (IRRID): RR2-10.2196/24647

(J Med Internet Res 2022;24(5):e34853) doi:10.2196/34853

KEYWORDS

sexual minority women; alcohol; intervention; social norms; gamification; mobile phone; smartphone

Introduction

Background

Relative to women who identify as heterosexual, experience only opposite-sex attractions, and only have sex with men, research has reliably documented more frequent and intense alcohol consumption [1-3], as well as a greater likelihood of negative alcohol-related consequences and alcohol dependence [2-4], among sexual minority women, a population that includes women who psychologically identify as lesbian, bisexual, or queer (LBQ), in addition to those who report having sex with women and experiencing same-sex attraction [5,6]. Although several culturally tailored interventions are currently being developed to meet the needs of heavy drinking sexual minority women seeking treatment for alcohol use disorder (AUD) [7,8], to date, there remains a lack of preventive, culturally tailored interventions to reduce alcohol-related harm among alcohol-consuming sexual minority women *who are not yet* motivated to reduce their drinking. Seeking to address this void, this study evaluates the degree to which an evidence-based personalized normative feedback (PNF) intervention embedded within a culturally tailored digital competition can engage LBQ-identified sexual minority women and reduce their alcohol-related risks.

Antecedents to Heavy Drinking Among Sexual Minority Women and Targets for Intervention

Consistent with the sexual minority stress model [9], extant research has linked greater alcohol consumption and negative consequences among sexual minority women to the internalization of sexual minority stigma [10-12] and experiences of harassment, discrimination, and violence due to sexual minority status [13-15]. These findings have informed the recent development of 2 stigma-coping–focused digital programs designed for heavy drinking sexual minority women seeking treatment for AUD only [7] and both AUD and poor mental health [8]. Although these programs hold promise for sexual minority women motivated to seek help, they appear unlikely to attract or engage the larger population of sexual minority women who do not view their mental health or drinking as problematic.

Recent research suggests that to motivate reductions in drinking among those not seeking treatment, it may be beneficial to target the elevated perceptions of sexual identity–specific drinking norms [16-20], which appear to be a consequence of the central role that alcohol use plays in queer socialization contexts [21-24]. Indeed, qualitative accounts from LBQ-identified women suggest that the position of bars and nightclubs as central hubs for queer socialization may lead young LBQ women to view heavy drinking as a normative *rite of passage* [23,24]. Findings from survey studies also suggest that the high visibility of alcohol use in physical and web-based LBQ community spaces may lead LBQ women to perceive heavy drinking as more characteristic or *typical* of LBQ peers than heterosexual women [20]. They tend to substantially overestimate how much and how often LBQ peers drink [16,17,19] and the frequency with which they drink to cope with sexual minority stigma [25].

Web-Based PNF Interventions

In other heavy drinking populations found to overestimate peer-drinking norms, alcohol-related risks have been reduced through PNF, a brief intervention strategy that only requires members of a social group to answer survey questions about their perceptions of the *typical* group member's drinking and then report on their own consumption [26-28]. Group members then receive individualized graphical reports highlighting discrepancies between their perceptions of peers' drinking, peers' actual drinking, and their own drinking [28,29]. To date, research has yet to investigate whether delivering PNF on LBQ-specific drinking and coping norms is an effective means of reducing alcohol-related risks among LBQ drinkers. However, supporting the promise of PNF for this population, in university and military samples, this strategy has been found particularly effective in reducing alcohol consumption among women [30,31], individuals for whom the reference peer group or community is important to their overall sense of self [32], those reporting coping motivations for drinking [33,34], and heavy drinkers not yet aware that their consumption exceeds normative standards [35].

Reaching LBQ Drinkers With PNF on Alcohol Use and Stigma Coping

Despite the potential promise of PNF, previous work suggests that LBQ women may comprise a population that is particularly difficult to reach, recruit, and retain in transparent health interventions. For instance, a review of community-based interventions targeting various health risk behaviors in this population identified low response rates, small sample sizes, and problems with attrition as significant challenges to evaluation efforts, reflecting broader difficulties with intervention engagement [36]. Recruitment and engagement concerns are also magnified in the PNF context, as this strategy is most effective in reducing alcohol-related risks among individuals who do not view their drinking as excessive or see themselves as in need of intervention. Moreover, very few PNF interventions have been delivered to populations not attached to institutions or workplaces, and researchers have struggled to implement PNF interventions outside study settings where participation is mandatory or participants are promised compensation at the point of recruitment [37-39].

Seeking to remedy these implementation challenges and extend promising gamified intervention work with college students [40-42], PNF on alcohol use and stigma-coping behaviors was

XSL•FO
RenderX

delivered to LBQ drinkers within *LezParlay*, a culturally tailored digital competition designed to challenge negative stereotypes about LBQ women and increase visibility (Figure 1). In brief, the competition comprised 8 monthly rounds wherein LBQ users guessed about the behaviors, attitudes, and experiences of age group and sexual identity–matched peers; wagered points on their guesses being true based on the responses of other users; and reported on their own corresponding behaviors, attitudes, and experiences. At the end of each month, players were SMS text messaged private URLs at which they could view detailed results (ie, PNF) on all or a subset of the round's questions. All actual norms presented in the detailed results (ie, PNF) were transparently derived from the responses of the players in each subgroup. Users' scores reflected the accuracy of their LBQ peer perceptions, and each round's top scorer won a variable cash prize. A complete overview of the digital competition and detailed descriptions of the theory-informed game mechanics and deep-structure cultural adaptations leveraged to bolster appeal and engagement are available in this project's protocol paper [43].

Figure 1. The initial version of LezParlay tested in this trial was a device-responsive HTML5 web application that delivered personalized normative feedback on a number of lesbian, bisexual, and queer stereotypes and health-related topics within the context of a monthly competition.



This Study

Informed by the Accelerated Creation-to-Sustainment model for the rapid development and evaluation of real-world–ready digital health interventions [44-46], a registered hybrid trial [43] simultaneously examined the degree to which LBQ women would be engaged in the LezParlay competition in the real world (when there was no study framing and participation incentives were not offered) and evaluated whether delivering PNF on alcohol use and stigma coping within the competition would meaningfully reduce alcohol use and negative consequences among participating LBQ drinkers. As shown in Textbox 1, this study examined 5 key questions related to LezParlay's feasibility and efficacy as an alcohol intervention strategy.

Textbox 1. Key feasibility and efficacy questions addressed in this study.

Feasibility

1. Were lesbian, bisexual, and queer (LBQ) women engaged by the LezParlay competition in the absence of traditional study incentives?
2. Could risk-reducing actual drinking norms be generated in real time from users' responses to round questions?
3. Did LBQ drinkers taking part in LezParlay find the competition acceptable and psychologically beneficial? What ideas for improvements were submitted?

Efficacy

1. Did personalized normative feedback (PNF) designed to correct LBQ peer-drinking norms reduce alcohol-related risks among LBQ drinkers?
2. Did PNF on both LBQ peer-drinking and stigma-coping norms better reduce alcohol-related risks than PNF on LBQ peer-drinking norms alone?

XSL•FO
RenderX

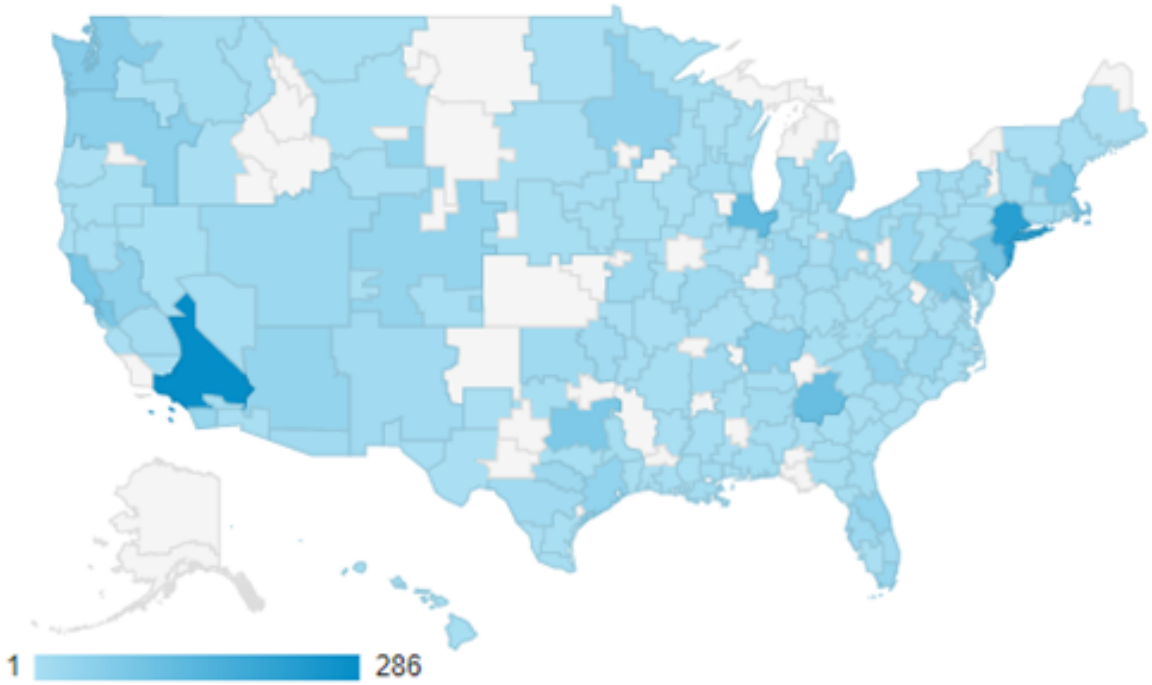
Methods

Participants and Procedure

Broader Competition

LezParlay was advertised to LBQ women as it would be in the real world—as a free, web-based competition designed to test LBQ stereotypes and increase visibility. Despite no traditional study incentives being offered at the point of recruitment or sign-up, 2677 LBQ women took part in the competition between December 2018 and July 2019 and played ≥1 of the 8 monthly rounds. LezParlay's informational landing page received 4099 unique views during recruitment and competition periods, with digital advertising campaigns responsible for the bulk of these views. Specifically, promotional campaigns on the HER Social app, a popular dating and social networking app for LBQ women, were responsible for 34.01% (1394/4099) of the total landing page visits, whereas campaigns on Facebook or Instagram and Google Search accounted for 32.01% (1312/4099) and 22.98% (942/4099) of the total visits, respectively. Of the 4099 landing page visitors, 2008 (48.99%) advanced directly to create a user account [43] on the LezParlay competition web app. In addition, 669 user accounts were created organically by users who did not view the landing page first but were directly invited to the LezParlay web app by a friend taking part in the competition. Figure 2 provides a visual breakdown of LezParlay users by US metropolitan area, and Table 1 presents basic user characteristics.

Figure 2. Geodensity of LezParlay users across US metropolitan areas.



XSL•FO
RenderX

Table 1. Characteristics of LezParlay users (N=2667).

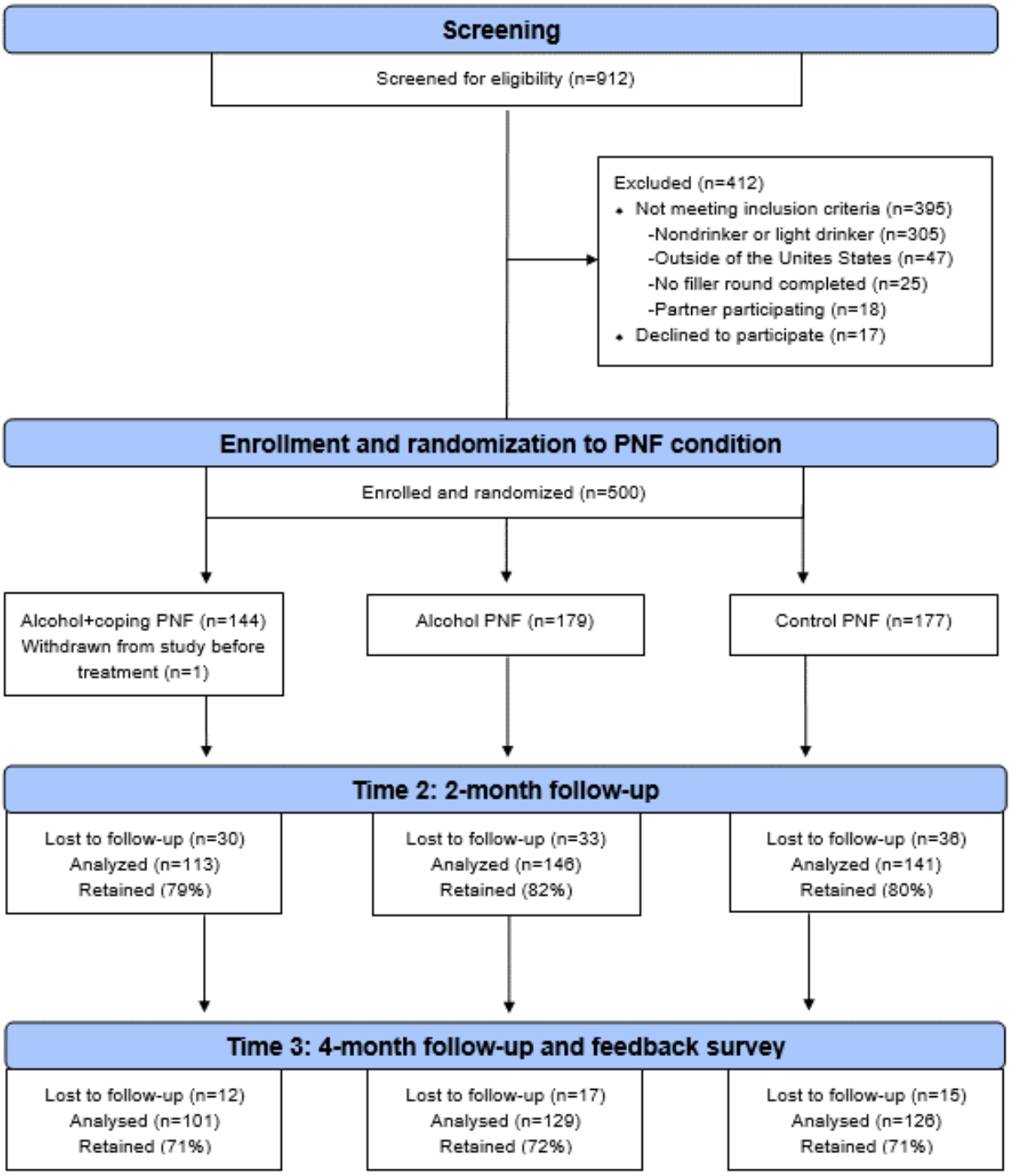
Characteristic	Participants, n (%)
Sexual identity	
Lesbian	1446 (54.22)
Bisexual	669 (25.08)
Queer	562 (21.07)
Age group (years)	
<18	107 (4.01)
18-24	401 (15.04)
25-34	1284 (48.14)
35-44	562 (21.07)
54-65	240 (9)
≥66	80 (3)
Relationship status	
Single	1205 (45.18)
In a relationship	857 (32.13)
Married	455 (17.06)
It's complicated	161 (6.04)
Device used	
Mobile phone	2266 (84.96)
Tablet	54 (2.02)
Computer	347 (13.01)

Evaluation Study (Randomized Controlled Trial)

The third monthly round of LezParlay inquired about alcohol use and LBQ stigma exposure and served as the screening instrument and baseline assessment (time point [T1]) for the randomized controlled trial (RCT). As players completed this round, a subsample of 500 LBQ drinkers meeting the eligibility criteria were invited to take part in an evaluation study wherein they were incentivized to play subsequent rounds and complete a feedback survey following the competition. A total of 1337 LBQ women completed the round with 912 users covertly screened for eligibility based on their responses to round questions about alcohol use (ie, reporting drinking ≥3 days per week or consuming ≥3 drinks on their heaviest drinking occasion) and other app data (eg, geolocation in the United States, at least one previous round played, and no partner participating) before the 500 spots in the evaluation study were filled. As described in greater detail elsewhere [43], at the point of study enrollment, participants were covertly randomized to receive 1 of 3 sequences of PNF delivered at the end of the third and fourth monthly rounds: alcohol + coping, alcohol + control, or control topics only. Reductions in drinking and negative consequences were assessed 2 (time point 2 [T2], June 2019) and 4 (time point 3 [T3], August 2019) months later. Following completion of a postcompetition feedback survey, participants were debriefed regarding the research questions and the nonrandom nature of the topics on which they received detailed results in 2 of the 8 competition rounds. A CONSORT (Consolidated Standards of Reporting Trials) diagram summarizing the flow of participants through the RCT portion of the trial is presented in Figure 3 (see also Multimedia Appendix 1 for this trial's CONSORT E-HEALTH checklist). Mirroring the larger user base, the evaluation study drinkers were diverse in terms of their geographic locations, representing 44 US states and 221 different counties and age groups, sexual identities, races, and ethnicities.

XSL•FO
RenderX

Figure 3. Flow of the evaluation showing study participants through screening, enrollment, and follow-ups. PNF: personalized normative feedback. T1: time point 1; T2: time point 2; T3: time point 3.



Ethical Considerations

Human subjects approval for this research was granted by the Loyola Marymount University institutional review board (protocol #LMUIRB2018SU14) on August 14, 2018. All procedures [43] were in accordance with the ethical standards of the US Department of Health and Human Services Office for Human Research.

Measures

Competition Engagement

Data collected by Google Analytics and the application examined the total number of users who signed up to participate in the LezParlay competition and detailed users' average number of logins, number of rounds completed, and most visited areas of the app.

XSL•FO
RenderX

Demographics

All users reported their sexual identity and relationship status at competition sign-up. Age in years, race, and ethnicity were also reported by the evaluation study participants at the point of study enrollment.

Perceived Norms for Alcohol Use and Negative Consequences

Three items modeled after the Quantity, Frequency, Max measure [47] assessed perceived LBQ peer alcohol use norms for the average number of drinks consumed in a typical week (2 items), maximum (or peak) number of drinks consumed on one occasion (1 item), and the number of negative alcohol-related consequences experienced (1 item). The perceived norm for weekly drinks comprised 2 items that prompted users to report their perception of the number of days per week the typical user drank (0-7 days) and the typical user's average number of drinks consumed per drinking day (0 to ≥12 drinks). The product of these 2 items was computed to create a variable indicative of users' perceived norm for weekly drinks. For the peak drinking norm, users reported the typical user's maximum number of drinks consumed on one occasion (0 to ≥12 drinks). To assess the perceived norm for negative consequences, users were presented with a list of 8 negative consequences (eg, had a hangover or illness, got in a physical or verbal fight, had problems with a significant other, missed a social engagement or event, had problems with friends or family, performed poorly at work or school, had problems with money, and had an unwanted or regrettable sexual experience) and were then asked to report the number they thought the typical user experienced due to drinking or partying. At all 3 time points, items assessing norms referenced the previous 2 months, and the sexual identity and age group of the typical user in these questions were piped to match each user's own sexual identity and age group (eg, "Over the past two months, on average, on how many days per week did the typical [lesbian] user in her [30s] drink?").

Own Alcohol Use and Negative Consequences

Users' own weekly drinks, peak drinks on one occasion, and negative consequences in the past 2 months were assessed using items that paralleled norm items (eg, "Over the past two months, on average, on how many days per week did YOU drink?") at the same time points and presented the same response options (ie, 0-7 days; 0 to ≥12 drinks; 0-8 negative consequences).

Interpersonal Stigma Exposure

Interpersonal stigma exposure was assessed at T1 with two items adapted for the in-game context from a widely used measure of sexual minority stigma [48]: (1) "During the past 2 months, how many times have you been physically harmed due to your sexual identity?" (2) "During the past 2 months, how many times have you been verbally harassed or threatened (online or in person) due to your sexual identity?" More than 75% of users' responses were concentrated in the range of 0 to 1. Therefore, items were first recoded to reflect this binary (ie, 0=this did not happen; 1=this happened once or more times) and then summed to produce a score between 0 and 2. This measure was included as a covariate in statistical models evaluating efficacy due to sexual minority status–related violence and harassment being the experiences most consistently linked to alcohol consumption and negative alcohol-related consequences among sexual minority women [13-15].

LezParlay Acceptability

The postcompetition feedback survey prompted study participants to rate numerous aspects of the competition (the stereotype challenge concept, topics and questions, detailed results, leaderboards, the ability to browse player profiles, the ability to submit questions, the ability to bet points on the accuracy of guesses, SMS text messages, and email communications) on Likert-type scales ranging from *did not like at all* (rating=0) to *liked very much* (rating=5). Total acceptability scores were computed by summing the ratings.

LezParlay Perceived Benefits

A single item in the feedback survey asked evaluation study participants to select yes or no in response to the question, "Did you find taking part in LezParlay to be psychologically beneficial?" Those selecting *yes* in response were invited to enter text describing their perceived benefits.

LezParlay Ideas for Improvement

A final free-response question asked evaluation study participants to share any ideas they had for ways in which LezParlay could be improved (ie, "Do you have any ideas for how LezParlay could be improved? What would you like to see in the next version?").

Analysis Plan

Evaluating Feasibility

Descriptive statistics examined the level of engagement with the LezParlay app (ie, total number of sign-ups and average number of rounds played), initial levels of alcohol use among players, competition-derived actual norms for treatment topics, competition acceptability among drinkers taking part in the RCT, and the proportion of drinkers taking part in the RCT reporting perceived psychological benefits. Qualitative responses to items assessing the perceived benefits associated with LezParlay and ideas for improvement were examined using a generic inductive qualitative coding approach [49,50]. This iterative approach is similar to grounded theory but is more commonly used in the program evaluation literature, where the coding objective tends to be summarizing phenomena for basic understanding rather than building theory. As this approach can be prone to reflecting the biases of a single coder, multiple coder perspectives were sought, with a lesbian-identified senior researcher (SCB), a gay male–identified senior researcher (JWL), and 2 heterosexual female research assistants taking part in the coding process. For psychological benefits, coding sought to condense the raw text descriptions submitted by participants into a summary of common benefit categories. First, the 2 senior researchers (SCB and JWL) conducted independent, initial readings of participant responses, with each aiming to identify no more than 10 unique categories of benefits. As responses were generally short, although many described >1 benefit, it was decided a priori that each response could receive up to 3 category classifications. The senior researchers then

XSL•FO
RenderX

met, compared categories, agreed upon common category themes, and identified several pairs of categories that were extremely similar and could be condensed into a single category. This process culminated in 6 shared benefit categories. Next, 2 research assistants independently classified all responses according to the 6 categories, with each response coded for a maximum of 3 benefits. Interrater reliability was high (κ=0.91), and discrepancies were resolved through discussion. A similar approach was used to code participants' ideas for improving LezParlay. However, as the research team only sought to identify the most agreed-upon or frequently submitted ideas for improvement to inform the next version of the app, all 4 coders independently grouped participant responses in terms of similarity. No discrepancies in groupings were observed between coders, and similar responses were tallied for summary purposes.

Evaluating Efficacy

Preliminary analysis of RCT data examined the distributions of alcohol-related variables, the nature of missing data, and the characteristics of participants lost to follow-up (*t* tests). One-way ANOVAs and chi-square tests also established conditional equivalency for demographic characteristics, interpersonal stigma, perceived alcohol use norms, and alcohol use behaviors assessed at T1. Examination of attrition suggested that missingness was random rather than systematic. As such, 3 multilevel models, each with maximum likelihood estimation to deal with data missing at random, a random intercept component, and an unstructured covariance matrix, were conducted in SAS (version 9.4) to assess the effects of treatment PNF (alcohol PNF and alcohol + coping PNF) on respective changes in weekly drinks, peak drinks, and negative consequences relative to control PNF. In all 3 models, predictors included study condition (alcohol PNF and alcohol + coping PNF vs control PNF) and time (T2 and T3 vs T1). To determine whether there were changes in drinking outcomes over time related to PNF treatment, condition × time interaction terms were included in each model. Covariates also included in the models were age, sexual identity (bisexual and queer vs lesbian), race (White vs non-White), ethnicity (non-Hispanic or Latino vs Hispanic or Latino), relationship status (single vs in a relationship or married), and exposure to interpersonal LBQ stigma. Post hoc Tukey tests were then conducted to determine the nature of significant condition × time effects.

Results

Research Question 1: Were LBQ Women Engaged by the LezParlay Competition?

Yes, despite no traditional study incentives being offered at recruitment, sign-up, or initial round completion, 2667 LBQ women signed up and played ≥1 round. Furthermore, the average user logged into the LezParlay app 2 times during the competition following initial sign-up; completed 1.97 rounds; and, on average, spent 4.15 minutes on the app per login (no SDs available). The LezParlay web application also recorded 54,072 total page visits among logged in users, with the most frequented sections of the app devoted to browsing the social media–inspired profiles of other users, followed by playing monthly rounds, viewing detailed results (ie, PNF), and viewing leaderboards.

Research Question 2: Was It Feasible to Derive Risk-Reducing LBQ Actual Drinking Norms From In-Round Questions?

Yes, of the 1337 LezParlay users who completed the round where alcohol use was first assessed, 254 (19%) reported no alcohol consumption or light drinking (≤2 drinks per week), and 346 (25.88%) reported moderate drinking (3-7 drinks per week and ≤3 drinks on any day). Notably, ≤7 drinks per week and ≤3 drinks per day are the upper limits for low-risk drinking among women, as defined by the National Institute on Alcohol Abuse and Alcoholism [51], as these patterns of consumption equate to low risks for alcohol dependence and development of alcohol-related health problems. Higher levels of risk were also well represented in LezParlay, with 55.12% (737/1337) of the users who completed this round consuming ≥8 drinks per week or ≥4 drinks on any day, thereby meeting the National Institute on Alcohol Abuse and Alcoholism's definition of high-risk drinking [51]. Conferring elevated risks for AUD and alcohol-related health problems, the average number of drinks consumed per week among these users ranged from 8 to 56 drinks, and peak drinks consumed on a day ranged from 4 to ≥12 drinks for the 2-month period referenced in the game questions. However, as LezParlay users were so diverse in their patterns of alcohol consumption, the broader composition of alcohol use among users was sufficient for generating risk-reducing actual norms to deliver to drinkers in the evaluation study. As is typically the case in traditional PNF interventions, the lower levels of consumption among nondrinkers and low-risk drinkers attenuated the higher levels of consumption among high-risk users. Round-derived actual norms featured in treatment PNF are presented in Table 2.

XSL•FO
RenderX

Table 2. Competition-derived actual norms presented in treatment personalized normative feedback.

Norms	Age groups (years)		
	21-29	30-39	≥40
Round 3: alcohol use actual norms[a]	n=627	n=498	n=212
Drinking days per week, mean	2	2	1.5
Drinks per occasion, mean	2.5	2	2
Weekly drinks, mean	5	4	3
Peak drinks on one occasion, mean	4	3	3
Negative consequences, mean	2	1.5	1
Round 4: coping actual norms[a]	n=503	n=414	n=186
Times drank alcohol to cope, %	18	17	16
Time used drugs to cope, %	12	9	9
Times exercised or meditated to cope, %	55	61	49
Times sought social support to cope, %	53	50	62

[a]As no sexual identity differences were observed for alcohol use or coping behaviors, participants received the same age group–specific lesbian, bisexual, and queer actual norms for these topics as a function of condition assignment.

Did PNF on Alcohol Use Delivered Within the Competition Reduce Alcohol-Related Risks Among LBQ Drinkers? Was It More Beneficial to Deliver PNF on Both Drinking and Coping Behaviors Than on Drinking Behaviors Alone?

Retention in the RCT was adequate, with 80.2% (400/499) of the evaluation study participants retained at T2 and 71.3% (356/499) at T3. Participants lost to follow-up were younger (t_{497}=4.48; *P*<.001) and non-Hispanic White (t_{497}=4.13; *P*<.001). Attrition, in this case, was considered random rather than systematic, given that the *average* participant in the study was both younger and non-Hispanic White, and attrition was not significantly associated with any other study variables. Beyond attrition, there were no other missing data among participants. As shown in Table 3, tests of conditional equivalency revealed no significant between-condition differences for any of the variables at baseline.

The results for multilevel models, predicting weekly drinks, peak drinks on one occasion, and negative alcohol-related consequences are presented in Table 4. The condition × time effects in each model were significant, indicating that treatment PNF conditions predicted significant changes in outcomes over time, controlling for baseline covariates (ie, sexual identity, race, ethnicity, age, relationship status, and interpersonal stigma exposure).

As presented in Table 5 and Figure 4, post hoc analyses probing interaction effects for each outcome revealed that participants in both treatment PNF conditions significantly decreased their weekly drinks from T1 to T2 relative to participants receiving PNF on control topics, but significant differences in weekly drinks were retained only at T3 between the alcohol + coping PNF and control PNF conditions. Similarly, both treatment PNF conditions predicted significant decreases in peak drinks consumed from T1 to T2 relative to control PNF; however, only the differences between the alcohol + coping and control PNF conditions met the threshold for significance at T3 (*P*=.06 for alcohol PNF vs control PNF). Finally, participants in both treatment PNF conditions significantly decreased the negative consequences they experienced relative to controls from T1 to T2 and from T2 to T3.

XSL•FO
RenderX

Table 3. Baseline demographics, psychosocial characteristics, drinking norms, and alcohol use of evaluation study participants overall and by condition assignment.

Characteristics	Overall (N=499)	Control PNF[a] (n=177)	Alcohol PNF (n=179)	Alcohol + coping PNF (n=143)
Sexual identity, n (%)				
Lesbian	290 (58.1)	94 (53.1)	108 (60.3)	89 (62.2)
Bisexual	115 (23)	48 (27.1)	39 (21.8)	29 (20.2)
Queer	94 (18.8)	35 (19.8)	32 (17.9)	25 (17.5)
Relationship status, n (%)				
Single	209 (41.9)	80 (45.2)	69 (38.5)	60 (41.9)
Ethnicity, n (%)				
Hispanic/Latino	123 (24.6)	40 (22.6)	46 (25.7)	37 (25.8)
Race, n (%)				
American Indian/Alaskan Native	13 (2.6)	4 (2.3)	5 (2.8)	4 (2.7)
Asian American	39 (7.8)	17 (9.6)	16 (8.9)	6 (4.1)
Black/African American	70 (14)	26 (14.7)	25 (14)	19 (13.2)
Hawaiian/Pacific Islander	1 (0.2)	0 (0)	0 (0)	1 (0.6)
White	268 (53.7)	99 (55.9)	91 (50.8)	78 (54.5)
Multiracial	53 (10.6)	15 (8.5)	23 (12.8)	15 (10.4)
Other	55 (11)	16 (9)	19 (10.6)	14 (20)
Age (years), mean (SD)	29.87 (7.32)	29.47 (7.03)	30.37 (7.75)	29.73 (7.15)
T1[b] interpersonal stigma, mean (SD)	0.61 (0.69)	0.66 (0.69)	0.56 (0.66)	0.62 (0.70)
T1 perceived drinking norms, mean (SD)				
Norm–weekly drinks	13.94 (9.37)	13.84 (9.44)	14.07 (10.35)	13.89 (7.92)
Norm–peak drinks	6.31 (2.18)	6.16 (2.27)	6.36 (2.07)	6.43 (2.20)
Norm–consequences	2.88 (1.74)	2.84 (1.65)	3.01 (1.84)	2.73 (1.72)
T1 alcohol use, mean (SD)				
Weekly drinks	9.15 (7.51)	9.13 (7.90)	8.96 (8.19)	9.43 (6.00)
Peak drinks	5.79 (2.34)	5.74 (2.43)	5.76 (2.37)	5.87 (2.19)
Consequences	2.52 (1.89)	2.45 (1.95)	2.55 (1.86)	2.58 (1.89)

[a]PNF: personalized normative feedback.

[b]T1: time point 1.

XSL•FO
RenderX

Table 4. Multilevel model results for outcomes (weekly drinks, peak drinks, and negative alcohol-related consequences).

Outcomes	Weekly drinks		Peak drinks		Consequences	
	b (SE)	*P* value	b (SE)	*P* value	b (SE)	*P* value
Alcohol PNF[a]	0.41 (0.74)	.58	0.13 (0.22)	.54	0.2 (0.16)	.24
Alcohol + coping PNF (reference: control PNF)	0.14 (0.78)	.86	0.16 (0.23)	.49	0.13 (0.17)	.47
Time 2	0.65 (0.39)	.09	0.19 (0.14)	.19	0.71 (0.12)	<.001
Time 3 (reference: time 1)	−0.50 (0.40)	.22	−0.05 (0.15)	.73	0.64 (0.13)	<.001
Alcohol PNF × time 2	−2.72 (0.54)	<.001	−1.61 (0.20)	<.001	−1.03 (0.18)	<.001
Alcohol PNF × time 3	−1.64 (0.57)	.004	−0.59 (0.21)	.005	−0.90 (0.18)	<.001
Alcohol + coping PNF × time 2	−3.39 (0.58)	<.001	−1.67 (0.22)	<.001	−1.00 (0.19)	<.001
Alcohol + coping PNF × time 3	−2.03 (0.61)	.01	−0.71 (0.23)	.02	−0.98 (0.20)	<.001
Queer	−1.88 (0.78)	.02	−0.29 (0.22)	.19	−0.48 (0.16)	.01
Bisexual (reference: lesbian)	−1.26 (0.72)	.08	0.40 (0.20)	.05	−0.03 (0.15)	.83
Non-White (reference: White)	0.75 (0.65)	.25	0.23 (0.18)	.21	0.51 (0.13)	.001
Hispanic or Latinx (reference: non-Hispanic or Latinx)	−0.68 (0.77)	.38	0.02 (0.22)	.93	−0.28 (0.15)	.07
Age	−0.05 (0.04)	.20	−0.04 (0.01)	.001	−0.04 (0.01)	<.001
Single (reference: coupled or married)	2.30 (0.62)	.002	0.61 (0.18)	.01	0.41 (0.13)	<.001
Interpersonal stigma exposure	1.85 (0.43)	<.001	0.38 (0.12)	.002	0.45 (0.09)	<.001

[a]PNF: personalized normative feedback.

Table 5. Tukey post hoc test results probing PNF[a] condition × time interactions.

PNF condition comparisons	Weekly drinks		Peak drinks		Consequences	
	b (SE)	*P* value	b (SE)	*P* value	b (SE)	*P* value
T1[b]						
Alcohol vs control	−0.41 (0.74)	.58	−0.13 (0.22)	.54	−0.20 (0.16)	.24
Alcohol + coping vs control	−0.14 (0.78)	.86	−0.16 (0.23)	.49	−0.13 (0.17)	.47
Alcohol vs alcohol + coping	0.27 (0.78)	.73	−0.03 (0.23)	.91	0.07 (0.17)	.69
T2[c]						
Alcohol vs control[d]	2.31 (0.77)	.003	1.48 (0.23)	<.001	0.83 (0.18)	<.001
Alcohol + coping vs control[d]	3.25 (0.82)	<.001	1.51 (0.25)	<.001	0.86 (0.19)	<.001
Alcohol vs alcohol + coping	0.94 (0.82)	.25	0.04 (0.25)	.88	0.03 (0.19)	.87
T3[e]						
Alcohol vs control[d]	1.24 (0.79)	.12	0.46 (0.24)	.06	0.70 (0.19)	.002
Alcohol + coping vs control[d]	1.90 (0.84)	.03	0.55 (0.26)	.03	0.86 (0.20)	<.001
Alcohol vs alcohol + coping	0.66 (0.84)	.43	0.10 (0.26)	.71	0.15 (0.20)	.44

[a]PNF: personalized normative feedback.

[b]T1: time point 1.

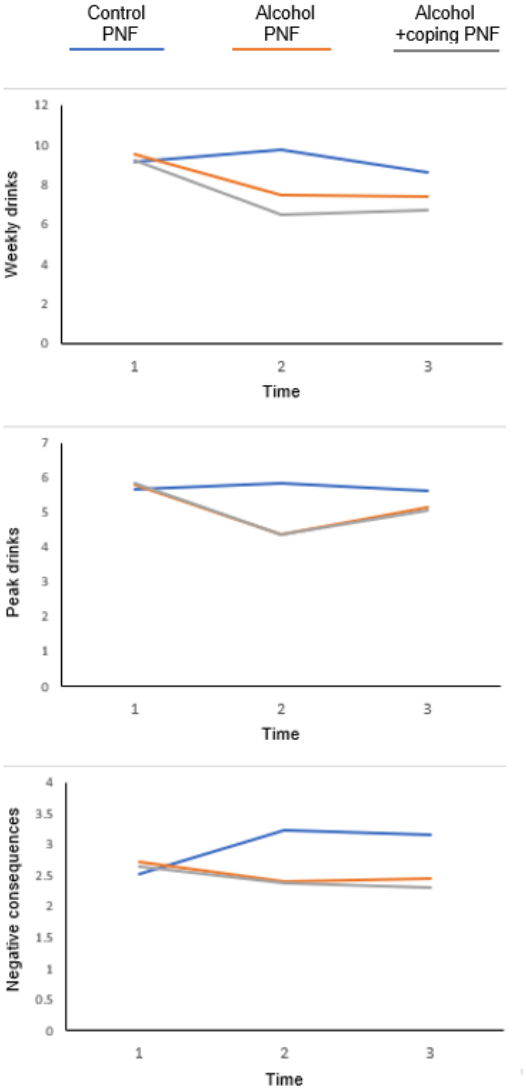
[c]T2: time point 2.

[d]Across outcomes, Cohen *d* effect size estimates for significant treatment versus control comparisons ranged from 0.20 to 0.33 at T2 and 0.12 to 0.22 at T3.

[e]T3: time point 3.

XSL•FO
RenderX

Figure 4. Personalized normative feedback (PNF) condition as a function of time for each outcome.



Did LBQ Drinkers Find the Competition Acceptable and Psychologically Beneficial? What Ideas for Improvements Did They Submit?

Overall, drinkers in the evaluation study found the LezParlay competition to be highly acceptable, with the average participant rating competition aspects between *liked* and *liked very much* (mean 41.26, SD 3.84; out of a maximum score of 50). Table 6 presents descriptive statistics for individual acceptability items. Notably, the highest-rated aspect was *receiving detailed results each round* (mean 4.51, SD 0.56). The exploratory 1-way ANOVA and correlational analyses also determined that acceptability ratings were not significantly associated with participants' study condition ($F_{2,355}$=0.41, P=.67) or baseline measures of alcohol outcomes (r ranged from 0.02 to 0.04; P ranged from .64 to .68).

Of the 356 participants who completed the feedback survey, 331 (93%) reported finding the LezParlay competition to be psychologically beneficial. Furthermore, 85.5% (283/331) of the participants that indicated benefits entered text to describe the experienced benefits. Qualitative coding resulted in 6 common categories of benefits reported by participants: knowledge and social comparison, community connection and identity strength, stigma reduction, introspection and self-confrontation, entertainment and fun, and mood enhancement. Table 7 presents the proportion of total responses that reflected each benefit category and representative responses of benefits in each category.

Notably, although benefits associated with social comparison and self-confrontation in the domain of drinking may be experienced in a traditional PNF alcohol intervention, benefits associated with community connection and identity strength, stigma reduction, entertainment, and enhanced mood and outlook would not be experienced in the context of traditional PNF. Presumably, these extra psychological benefits described by participants were related to LezParlay's social media–inspired web-based community features, the broad challenging of negative LBQ stereotypes via PNF, and the competition-fostering constellation of game mechanics.

In response to the optional item inquiring as to how LezParlay could be improved, 47.5% (169/356) of participants submitted a total of 307 individual ideas for improvement. The most frequently submitted ideas revealed that participants most commonly desired a native (iOS and Android) smartphone app

XSL•FO
RenderX

for the competition (112/307, 36.5%); more frequent rounds (eg, weekly rather than monthly) with faster results delivery (74/307, 24.1%); increased opportunity for interaction between players (eg, a chat feature or direct messaging that could be turned on and off; 43/307, 14%); increased ease of inviting friends and the ability to earn bonus points for referring friends (24/307, 7.8%); the ability to go back and change previously submitted guesses or point wagers before a round closing (12/307, 3.9%); the connection of results to informational articles or community resources (11/307, 3.6%); additional questions about race, gender identity, and sexual identity–based biases within the community (9/307, 2.9%); worldwide promotion and additional results comparing the behaviors and experiences of LBQ players in different countries or regions (4/307, 1.3%); and the ability to see the community thumbnail photo collage of LBQ users being guessed about on the *guess question* screen rather than on a previous screen (4/307, 1.3%).

Table 6. LezParlay competition acceptability ratings by item (N=356).

Acceptability item	Rating, mean (SD)[a]
The “stereotype challenge” concept	4.37 (0.56)
The topics and questions	4.01 (0.56)
Receiving the detailed results each round	4.51 (0.56)
Browsing players profiles	3.62 (0.66)
Submitting and voting on questions	3.95 (0.59)
Betting points on your guesses being correct	4.34 (0.62)
Receiving SMS text message reminders	4.12 (0.58)
Receiving email reminders	4.27 (0.50)
Viewing the top scorer leaderboards	3.89 (0.60)
Competing for money and receiving gift cards	4.09 (0.55)

[a]Response options ranged from 1=disliked very much to 5=liked very much.

Table 7. Categories of psychological benefits described by participants and representative responses (n=283).

Benefit category	Total, n (%)	Representative responses and user characteristics
Knowledge and social comparison	184 (65)	• “I work at an LGBTQ community center and it really helped having data to influence our programs and identify topics/issues to discuss in our women's group meetings.” [Queer, 41 years] • “Let me learn more about the lgbtq community and see that I drink way more than average lol fail” [Lesbian, 38 years]
Community connection and identity strength	96 (33.9)	• “Being in my 50s and feeling sort of invisible these days this competition really helped me feel connected to something again” [Lesbian, 52 years] • “It was so great for me although it's hard to describe how/why exactly...felt connected and in the know...also felt more confident and secure in my identity.” [Queer, 25 years]
Stigma reduction	88 (31)	• “This really helped me reduce biases that I had internalized without even realizing it!” [Queer, 26 years] • “Cool to see that some of the negative ways we get portrayed in the media are totally off.” [Lesbian, 36 years]
Introspection and self-confrontation	66 (23.3)	• “This really helped me see that I need to get my shit together in several areas” [Lesbian, 33 years] • “Made me question some of my own tendencies and behaviors. Came to see that I was doing what I thought everyone else was doing which wasn't even the truth...” [Lesbian, 28 years]
Funa and entertainment	54 (19)	• “Loved the competition, betting, prizes, and leaderboards... so so fun” [Lesbian, 51 years] • “It was so fun and I was able to get my lesbian roommate to play with me...we got super competitive about scores and had a blast!” [Bisexual, 37 years]
Mood and outlook enhancement	40 (14.1)	• “...helped my mental health and gave me a more positive outlook on all things queer.” [Queer, 29 years] • “It was a source of enjoyment. Getting the results always put me in such a good mood...even when I was wrong about stuff...” [Lesbian, 23 years]

XSL•FO
RenderX

Discussion

Principal Findings

LezParlay leveraged gamification and deep cultural adaptations to deliver a PNF alcohol intervention to LBQ women, a difficult-to-engage population for whom alcohol-related risks are high, but efficacious evidence-based intervention and prevention programs are lacking [3,52]. Reflecting the widespread appeal and cost-efficacy afforded by LezParlay's framing as a competition designed to challenge negative LBQ stereotypes, a very large and diverse group of LBQ women signed up to take part in the competition despite the lack of traditional study incentives being offered for sign-up or initial round play. Furthermore, more than half of the LBQ users who completed the round in which alcohol use was first assessed exceeded established drinking guidelines for women and thus were an ideal population for PNF intervention. LBQ women taking part in LezParlay substantially overestimated LBQ-specific peer norms for drinking, experiencing negative consequences, and engaging in maladaptive coping behaviors in response to stress and stigma consistent with previous survey study findings [16-19]. In summary, the markedly lower levels of consumption among alcohol abstainers and low-risk drinkers also taking part in the round attenuated the levels of consumption among heavier drinkers, allowing risk-reducing actual drinking norms (presented in PNF) to be organically generated in real time from users' round data.

This novel approach to PNF intervention also demonstrated efficacy in reducing drinking and its negative consequences. Relative to LBQ drinkers randomized to receive PNF on control topics in the competition, those who received treatment PNF on drinking or both drinking and coping similarly and substantially reduced their weekly drinks, peak drinks consumed on one occasion, and number of negative consequences 2 months later. For these outcomes, effect sizes associated with LezParlay treatment arms at the 2-month follow-up were consistent with or exceeded the short-term effects associated with treatment arms of traditional, remotely delivered PNF alcohol interventions in other populations [53-55]. Thus, in the short term, the impact of additional treatment PNF on coping behaviors beyond alcohol PNF was negligible. However, at the 4-month follow-up, relative to control PNF, the reductions in quantity of consumption outcomes (ie, weekly drinks and peak drinks) associated with the alcohol-only PNF condition faded, whereas they remained relatively robust in the alcohol + coping condition. There are 2 potential explanations for this finding. First, as previous research has found drinking to cope to be a strong overall predictor of alcohol consumption among LBQ women [56,57], correcting LBQ coping norms may have changed participants' own coping behaviors to be more adaptive, which, with passing time, impacted alcohol use outcomes. However, given the design of this trial, wherein coping PNF corrected norms for coping-motivated drinking (among other behaviors) and was delivered 1 month following the initial treatment PNF on alcohol use, it is possible that coping PNF had little effect on participants' subsequent coping behaviors. Rather, the portion of coping PNF that corrected norms for coping-motivated drinking may have acted as a broader *booster* to alcohol PNF, further reinforcing the idea that LBQ peers do not drink as much as one previously thought. Thus, although efficacy findings from this initial trial are promising and suggest that both alcohol and coping PNF are beneficial, additional research will be needed to fully understand the mechanisms (ie, correcting coping norms or reinforcing actual alcohol use norms) by which coping PNF influences drinking in this population.

In addition to being a feasible and effective means of delivering PNF to this population, LBQ drinkers also found the LezParlay competition to be both highly acceptable and psychologically beneficial. Notably, the detailed results (ie, PNF) were the most liked aspect of LezParlay, and ratings were not significantly associated with study condition or baseline alcohol consumption. Thus, those receiving fewer and more health-related results as a function of condition and those entering the study as lighter and heavier drinkers similarly enjoyed receiving the PNF delivered. These findings suggest that future versions of the competition might also correct additional types of alcohol and coping-related norms or expand the topics on which PNF is delivered to other areas of physical and mental health without detracting from acceptability or engagement. Most participants also reported that they psychologically benefited from taking part in the competition, and descriptions of benefits reflected learning and social comparison, community connection and identity strength, stigma reduction, enhanced mood or outlook, and entertainment. Many of these benefits map onto LezParlay's social media–inspired web-based community features, the constellation of game mechanics, and the broad challenging of negative LBQ stereotypes and, importantly, extend far beyond the psychological benefits associated with traditional PNF alcohol interventions. Finally, participants submitted several actionable ideas for ways in which the LezParlay app, user experience, and competition format could be improved. These insights will inform the next version of LezParlay.

Implications for Intervention Research and Practice

To date, alcohol interventions developed for sexual minorities have tended to be clinical, intensive, and focused on affirming sexual identities, aiding individuals in understanding sexual minority stress processes, and providing resources to help individuals cope with stigma more adaptively [52,58,59]. Although these approaches hold much promise for individuals seeking treatment, other findings suggest that the central and highly visible positions that bars and nightclubs occupy in sexual minority communities may diminish community members' recognition of their heavy drinking as problematic and motivation to change, thereby deterring or delaying treatment seeking [60-62]. However, very few, if any, previous evidence-based interventions have been designed to motivate reductions in drinking among sexual minorities who do not view their drinking as problematic or experience other barriers to intensive treatment programs. To our knowledge, this study is the first to demonstrate that correcting sexual minority–specific drinking and coping norms via PNF is effective in reducing drinking in a sexual minority population. Although more research is needed, these findings suggest that the impact of PNF is not diminished by violence and harassment due to sexual minority status and that this approach may similarly reduce alcohol-related risks in other populations of sexual minority

XSL•FO
RenderX

adolescents and adults. Importantly, for LBQ women, this gamified, incognito, brief intervention also provides a valuable complement to more intensive programs being developed to meet the needs of self-aware LBQ women already motivated to reduce their consumption [7] and those seeking culturally tailored treatment for AUD and comorbid mental health problems [8].

The present findings also bring clarity to confusion in the substance use literature around the appropriateness and utility of social norms interventions for health disparity populations [18]. For instance, because it is well-known that LBQ women disproportionately drink and experience consequences relative to heterosexual women, there is often confusion as to whether delivering PNF on LBQ-specific actual alcohol use norms would have the effect of increasing or reducing drinking. As evident from this trial's findings, increasing drinking should not be a concern with this approach to the extent that LBQ women overestimate the drinking of other LBQ women; that is, similar to college students known to drink disproportionately drink relative to noncollege students and military populations known to drink more than their civilian peers, PNF reduces alcohol use in these heavy drinking populations despite disparate out-group comparisons. This type of intervention is effective because perceived in-group drinking norms are both highly relevant to the self and substantially overestimated.

This study's feasibility findings also provide an innovative answer to challenges related to reaching and engaging stigmatized minority populations with PNF in the real world. LezParlay delivered the core components of a PNF intervention within a fun, culturally tailored digital competition designed to challenge negative stereotypes about the target population. This gamified, incognito intervention approach was found to be highly engaging, acceptable, and psychologically beneficial among alcohol-consuming LBQ women and meaningfully reduced their alcohol-related risks. Although more research is needed, the stereotype challenge concept, along with the injection of established game mechanics and cultural themes, may have similar utility in reaching other high-risk stigmatized minority groups with PNF on drinking and other health risk behaviors. Finally, looking past PNF, findings from this study also suggest that challenging negative identity-related stereotypes and including web-based community features may also prove fruitful in minority, stress-informed digital health and mental health programs targeting internalized stigma, loneliness, and isolation [8,63,64].

Limitations and Future Directions

As the initial step in a new direction for alcohol intervention development, the key limitations associated with this study include the relatively short duration of the follow-up period, organic assessment of baseline and follow-up alcohol outcomes within rounds of the competition at T1 and T2, and assessment of acceptability and psychological benefits only among LBQ drinkers involved in the RCT. Thus, future evaluation efforts should follow participants for a longer duration (6-24 months), incorporate survey-based baseline and follow-up assessments, and examine the acceptability and psychological benefits among nondrinkers and other LBQ players not involved in the efficacy portion of the trial. An additional limitation to be remedied in future research is this trial's lack of an assessment-only control condition. Although randomizing participants to receive PNF on either treatment or control topics, as was done in this study, reflects the gold standard trial design in the PNF intervention literature, it may not be optimal when PNF is delivered within a culturally tailored digital competition focused on challenging negative group stereotypes. That is, participants in all 3 PNF conditions described unanticipated, far-reaching psychological benefits associated with broader participation in the competition, including stigma reduction, community connection, and identity strength. As these factors are theorized to diminish the degree to which sexual minority stress negatively impacts health risk behavior [65,66], it is possible that they alone may have reduced drinking across PNF conditions to some degree. To fully determine the impact of the LezParlay competition app as an alcohol intervention strategy, it will be important for future trials to also include an assessment-only control group with no exposure to PNF or the competition app. Future trials using such an expanded design should examine internalized stigma, LBQ identity strength, and community connection, in addition to perceived norms for treatment topics as potential mediators of conditional effects on drinking and negative consequences. Similarly, it will also be important to examine internalized, structural, and interpersonal forms of sexual minority stigma as potential moderators of direct and indirect effects.

Although this efficacious initial version of LezParlay was a standalone intervention focused exclusively on correcting descriptive drinking and coping norms, exciting directions for future research also lie in the prospect of incorporating additional components to further reduce alcohol-related risks and increase wellness more broadly. For example, future research may seek to evaluate the utility of including a judgment-based reflective injunctive alcohol normative feedback component that builds on promising pilot findings among college students [41]. The competition's multiround format also provides a natural environment for examining the utility of PNF on dynamic or *trending* health-related norms [67,68] focused on group-based changes in behavior or attitudes over time. Furthermore, the competition's ability to attract and engage LBQ women in the absence of traditional study incentives also suggests that it could play a future role in implementing more intensive health interventions that have found it difficult to engage this population [36]. Thus, another important direction for future research is to examine the degree to which LezParlay could fruitfully serve to attract LBQ women and motivate behavior change as part of a larger multicomponent intervention targeting multiple health behaviors. For instance, within the competition, PNF could target additional health behaviors, and after motivating behavior change through norms correction, the app could link at-risk users to intensive web-based intervention components or local health promotion programs that correspond to these behaviors.

Conclusions

The results of this hybrid trial provide initial support for the feasibility and efficacy of LezParlay as a culturally tailored, gamified, PNF alcohol intervention for LBQ women, thereby narrowing costly disparities in alcohol intervention research

XSL•FO
RenderX

and practice. The reductions in alcohol use and negative consequences associated with PNF on drinking and coping delivered within LezParlay demonstrate the utility of PNF as an alcohol intervention strategy for a stigmatized minority health disparity population. These findings should behoove substance use researchers developing interventions for sexual minorities to consider such sexual identity–specific peer norms as potential intervention targets. Furthermore, to overcome engagement challenges associated with delivering PNF to non–treatment-seeking members of stigmatized minority groups and broaden the psychological benefits associated with this strategy, the findings underscore the value of packaging PNF within a broader culturally tailored competition designed to challenge negative group stereotypes.

Acknowledgments

Manuscript preparation was supported by developmental exploratory grant R21AA025767-01A from the National Institute on Alcohol Abuse and Alcoholism. The content is solely the responsibility of the authors and does not necessarily reflect the official views of the National Institute on Alcohol Abuse and Alcoholism or its affiliates.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 25450 KB - jmir_v24i5e34853_app1.pdf]

References

1. Fish JN. Sexual orientation-related disparities in high-intensity binge drinking: findings from a nationally representative sample. LGBT Health 2019 Jul;6(5):242-249 [FREE Full text] [doi: 10.1089/lgbt.2018.0244] [Medline: 31184966]
2. Hughes T. Alcohol-related problems among sexual minority women. Alcohol Treat Q 2011;29(4):403-435 [FREE Full text] [doi: 10.1080/07347324.2011.608336] [Medline: 22470226]
3. Hughes TL, Veldhuis CB, Drabble LA, Wilsnack SC. Research on alcohol and other drug (AOD) use among sexual minority women: a global scoping review. PLoS One 2020 Mar 18;15(3):e0229869 [FREE Full text] [doi: 10.1371/journal.pone.0229869] [Medline: 32187200]
4. McCabe SE, Hughes TL, Bostwick WB, West BT, Boyd CJ. Sexual orientation, substance use behaviors and substance dependence in the United States. Addiction 2009 Aug;104(8):1333-1345 [FREE Full text] [doi: 10.1111/j.1360-0443.2009.02596.x] [Medline: 19438839]
5. Sell RL. Defining and measuring sexual orientation: a review. Arch Sex Behav 1997 Dec;26(6):643-658. [doi: 10.1023/a:1024528427013] [Medline: 9415799]
6. Talley AE, Sher KJ, Steinley D, Wood PK, Littlefield AK. Patterns of alcohol use and consequences among empirically derived sexual minority subgroups. J Stud Alcohol Drugs 2012 Mar;73(2):290-302 [FREE Full text] [doi: 10.15288/jsad.2012.73.290] [Medline: 22333337]
7. Bush R, Brown R, McNair R, Orellana L, Lubman DI, Staiger PK. Effectiveness of a culturally tailored SMS alcohol intervention for same-sex attracted women: protocol for an RCT. BMC Womens Health 2019 Feb 06;19(1):29 [FREE Full text] [doi: 10.1186/s12905-019-0729-y] [Medline: 30728002]
8. Pachankis JE, McConocha EM, Clark KA, Wang K, Behari K, Fetzner BK, et al. A transdiagnostic minority stress intervention for gender diverse sexual minority women's depression, anxiety, and unhealthy alcohol use: a randomized controlled trial. J Consult Clin Psychol 2020 Jul;88(7):613-630 [FREE Full text] [doi: 10.1037/ccp0000508] [Medline: 32437174]
9. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. Psychol Bull 2003 Sep;129(5):674-697 [FREE Full text] [doi: 10.1037/0033-2909.129.5.674] [Medline: 12956539]
10. Bandermann KM, Szymanski DM. Exploring coping mediators between heterosexist oppression and posttraumatic stress symptoms among lesbian, gay, and bisexual persons. Psychol Sex Orientat Gend Divers 2014;1(3):213-224. [doi: 10.1037/sgd0000044]
11. Dorn-Medeiros CM, Doyle C. Alcohol as coping: internalized homophobia and heterosexism's role in alcohol use among lesbians. J LGBT Issues Couns 2018;12(3):142-157. [doi: 10.1080/15538605.2018.1488230]
12. Lehavot K, Simoni JM. The impact of minority stress on mental health and substance use among sexual minority women. J Consult Clin Psychol 2011 Apr;79(2):159-170 [FREE Full text] [doi: 10.1037/a0022839] [Medline: 21341888]
13. Lewis RJ, Winstead BA, Mason TB, Lau-Barraco C. Social factors linking stigma-related stress with alcohol use among lesbians. J Soc Issues 2017 Sep;73(3):545-562 [FREE Full text] [doi: 10.1111/josi.12231] [Medline: 29269960]

XSL•FO
RenderX

14. McNair R, Pennay A, Hughes T, Brown R, Leonard W, Lubman DI. A model for lesbian, bisexual and queer-related influences on alcohol consumption and implications for policy and practice. Cult Health Sex 2016;18(4):405-421. [doi: 10.1080/13691058.2015.1089602] [Medline: 26466746]
15. Wilson SM, Gilmore AK, Rhew IC, Hodge KA, Kaysen DL. Minority stress is longitudinally associated with alcohol-related problems among sexual minority women. Addict Behav 2016 Oct;61:80-83 [FREE Full text] [doi: 10.1016/j.addbeh.2016.05.017] [Medline: 27249806]
16. Boyle SC, LaBrie JW, Witkovic YD. Do lesbians overestimate alcohol use norms? Exploring the potential utility of personalized normative feedback interventions to reduce high-risk drinking in Southern California lesbian communities. J Gay Lesbian Soc Serv 2016;28(3):179-194 [FREE Full text] [doi: 10.1080/10538720.2016.1190677] [Medline: 28579731]
17. Boyle SC, Kettering V, Young SH, LaBrie JW. Alcohol Treat Q 2020;38(4):415-429 [FREE Full text] [doi: 10.1080/07347324.2020.1723459] [Medline: 34108797]
18. Boyle SC, LaBrie JW, Omoto AM. Normative substance use antecedents among sexual minorities: a scoping review and synthesis. Psychol Sex Orientat Gend Divers 2020 Jun;7(2):117-131 [FREE Full text] [doi: 10.1037/sgd0000373] [Medline: 34079845]
19. Ehlke SJ, Stamates AL, Kelley ML, Braitman AL. Bisexual women's reports of descriptive drinking norms for heterosexual, bisexual, and lesbian women. Psychol Sex Orientat Gend Divers 2019 Jun;6(2):256-263 [FREE Full text] [doi: 10.1037/sgd0000312] [Medline: 31106227]
20. Litt DM, Lewis MA, Rhew IC, Hodge KA, Kaysen DL. Reciprocal relationships over time between descriptive norms and alcohol use in young adult sexual minority women. Psychol Addict Behav 2015 Dec;29(4):885-893 [FREE Full text] [doi: 10.1037/adb0000122] [Medline: 26478944]
21. Cogger A, Conover KJ, Israel T. Factors influencing alcohol use among sexual minority women in a non-urban community: a mixed methods study. J LGBT Issues Couns 2012;6(4):293-309. [doi: 10.1080/15538605.2012.727745]
22. Drabble L, Trocki KF. Alcohol in the life narratives of women: commonalities and differences by sexual orientation. Addict Res Theory 2014 Jun 01;22(3):186-194 [FREE Full text] [doi: 10.3109/16066359.2013.806651] [Medline: 24955083]
23. Gruskin E, Byrne K, Kools S, Altschuler A. Consequences of frequenting the lesbian bar. Women Health 2006;44(2):103-120. [doi: 10.1300/j013v44n02_06] [Medline: 17255061]
24. Parks CA. Lesbian social drinking: the role of alcohol in growing up and living as lesbian. Contemp Drug Probl 1999;26(1):75-129 [FREE Full text] [doi: 10.1177/009145099902600105]
25. Boyle SC, LaBrie JW, Costine LD, Witkovic YD. "It's how we deal": perceptions of LGB peers' use of alcohol and other drugs to cope and sexual minority adults' own coping motivated substance use following the Pulse nightclub shooting. Addict Behav 2017 Feb;65:51-55 [FREE Full text] [doi: 10.1016/j.addbeh.2016.10.001] [Medline: 27728830]
26. Berkowitz AD. The social norms approach: theory, research, and annotated bibliography. 2004. URL: http://www.alanberkowitz.com/articles/social_norms.pdf [accessed 2011-01-21]
27. Perkins HW. The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians. San Francisco, CA, USA: Jossey-Bass; 2003.
28. Perkins HW, Berkowitz AD. Perceiving the community norms of alcohol use among students: some research implications for campus alcohol education programming. Int J Addict 1986;21(9-10):961-976. [doi: 10.3109/10826088609077249] [Medline: 3793315]
29. Miller DT, Prentice DA. Changing norms to change behavior. Annu Rev Psychol 2016;67:339-361. [doi: 10.1146/annurev-psych-010814-015013] [Medline: 26253542]
30. Murphy JG, Benson TA, Vuchinich RE, Deskins MM, Eakin D, Flood AM, et al. A comparison of personalized feedback for college student drinkers delivered with and without a motivational interview. J Stud Alcohol 2004 Mar;65(2):200-203. [doi: 10.15288/jsa.2004.65.200] [Medline: 15151350]
31. Saunders JB, Kypri K, Walters ST, Laforge RG, Larimer ME. Approaches to brief intervention for hazardous drinking in young people. Alcohol Clin Exp Res 2004 Feb;28(2):322-329. [doi: 10.1097/01.alc.0000113418.12889.08] [Medline: 15112940]
32. Neighbors C, LaBrie JW, Hummer JF, Lewis MA, Lee CM, Desai S, et al. Group identification as a moderator of the relationship between perceived social norms and alcohol consumption. Psychol Addict Behav 2010 Sep;24(3):522-528 [FREE Full text] [doi: 10.1037/a0019944] [Medline: 20853938]
33. Young CM, Neighbors C, DiBello AM, Sharp C, Zvolensky MJ, Lewis MA. Coping motives moderate efficacy of personalized normative feedback among heavy drinking U.S. college students. J Stud Alcohol Drugs 2016 May;77(3):495-499 [FREE Full text] [doi: 10.15288/jsad.2016.77.495] [Medline: 27172582]
34. Young CM, Pedersen ER, Pearson AD, Neighbors C. Drinking to cope moderates the efficacy of changing veteran drinking norms as a strategy for reducing drinking and alcohol-related problems among U.S. veterans. Psychol Addict Behav 2018 Mar;32(2):213-223 [FREE Full text] [doi: 10.1037/adb0000347] [Medline: 29369674]
35. Bertholet N, Daeppen JB, Cunningham JA, Burnand B, Gmel G, Gaume J. Are young men who overestimate drinking by others more likely to respond to an electronic normative feedback brief intervention for unhealthy alcohol use? Addict Behav 2016 Dec;63:97-101. [doi: 10.1016/j.addbeh.2016.07.015] [Medline: 27450907]

XSL•FO
RenderX

36. Rizer AM, Mauery DR, Haynes SG, Couser B, Gruman C. Challenges in intervention research for lesbian and bisexual women. LGBT Health 2015 Jun;2(2):105-112. [doi: 10.1089/lgbt.2014.0122] [Medline: 26790115]
37. Bedendo A, Ferri CP, de Souza AA, Andrade AL, Noto AR. Pragmatic randomized controlled trial of a web-based intervention for alcohol use among Brazilian college students: motivation as a moderating effect. Drug Alcohol Depend 2019 Jun 01;199:92-100. [doi: 10.1016/j.drugalcdep.2019.02.021] [Medline: 31029880]
38. Dempsey RC, McAlaney J, Bewick BM. A critical appraisal of the social norms approach as an interventional strategy for health-related behavior and attitude change. Front Psychol 2018 Nov 6;9:2180 [FREE Full text] [doi: 10.3389/fpsyg.2018.02180] [Medline: 30459694]
39. Palfai TP, Winter M, Lu J, Rosenbloom D, Saitz R. Personalized feedback as a universal prevention approach for college drinking: a randomized trial of an e-mail linked universal web-based alcohol intervention. J Prim Prev 2014 Apr;35(2):75-84 [FREE Full text] [doi: 10.1007/s10935-013-0337-9] [Medline: 24421075]
40. Boyle SC, Earle AM, LaBrie JW, Smith DJ. PNF 2.0? Initial evidence that gamification can increase the efficacy of brief, web-based personalized normative feedback alcohol interventions. Addict Behav 2017 Apr;67:8-17 [FREE Full text] [doi: 10.1016/j.addbeh.2016.11.024] [Medline: 27978426]
41. Earle AM, LaBrie JW, Boyle SC, Smith D. In pursuit of a self-sustaining college alcohol intervention: deploying gamified PNF in the real world. Addict Behav 2018 May;80:71-81 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.005] [Medline: 29407688]
42. LaBrie JW, de Rutte JL, Boyle SC, Tan CN, Earle AM. Leveraging copresence to increase the effectiveness of gamified personalized normative feedback. Addict Behav 2019 Dec;99:106085 [FREE Full text] [doi: 10.1016/j.addbeh.2019.106085] [Medline: 31421583]
43. Boyle SC, LaBrie JW. A gamified, social media-inspired, web-based personalized normative feedback alcohol intervention for lesbian, bisexual, and queer-identified women: protocol for a hybrid trial. JMIR Res Protoc 2021 Apr 16;10(4):e24647 [FREE Full text] [doi: 10.2196/24647] [Medline: 33861212]
44. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
45. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
46. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 08;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
47. Baer J. Etiology and secondary prevention of alcohol problems with young adults. In: Baer JS, Marlatt GA, McMahon RJ, editors. Addictive Behaviors Across the Life Span: Prevention, Treatment, and Policy Issues. Thousand Oaks, CA, USA: Sage Publications; 1993:111-137.
48. Herek GM. Hate crimes and stigma-related experiences among sexual minority adults in the United States: prevalence estimates from a national probability sample. J Interpers Violence 2009 Jan;24(1):54-74. [doi: 10.1177/0886260508316477] [Medline: 18391058]
49. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246. [doi: 10.1177/1098214005283748]
50. Liu L. Using generic inductive approach in qualitative educational research: a case study analysis. J Educ Learn 2016;5(2):129-135. [doi: 10.5539/jel.v5n2p129]
51. Alcohol Research: Current Reviews Editorial Staff. Drinking patterns and their definitions. Alcohol Res 2018;39(1):17-18 [FREE Full text] [Medline: 30557143]
52. Bloom AW. Advances in substance abuse prevention and treatment interventions among racial, ethnic, and sexual minority populations. Alcohol Res 2016;38(1):47-54 [FREE Full text] [Medline: 27159811]
53. Riper H, van Straten A, Keuken M, Smit F, Schippers G, Cuijpers P. Curbing problem drinking with personalized-feedback interventions: a meta-analysis. Am J Prev Med 2009 Mar;36(3):247-255. [doi: 10.1016/j.amepre.2008.10.016] [Medline: 19215850]
54. Dotson KB, Dunn ME, Bowers CA. Stand-alone personalized normative feedback for college student drinkers: a meta-analytic review, 2004 to 2014. PLoS One 2015 Oct 8;10(10):e0139518 [FREE Full text] [doi: 10.1371/journal.pone.0139518] [Medline: 26447792]
55. Doherty AM, Mason C, Fear NT, Rona R, Greenberg N, Goodwin L. Are brief alcohol interventions targeting alcohol use efficacious in military and veteran populations? A meta-analysis. Drug Alcohol Depend 2017 Sep 01;178:571-578. [doi: 10.1016/j.drugalcdep.2017.05.029] [Medline: 28750345]
56. Dworkin ER, Cadigan J, Hughes T, Lee C, Kaysen D. Sexual identity of drinking companions, drinking motives, and drinking behaviors among young sexual minority women: an analysis of daily data. Psychol Addict Behav 2018 Aug;32(5):540-551 [FREE Full text] [doi: 10.1037/adb0000384] [Medline: 30047754]

XSL•FO
RenderX

57. Lewis RJ, Mason TB, Winstead BA, Gaskins M, Irons LB. Pathways to hazardous drinking among racially and socioeconomically diverse lesbian women: sexual minority stress, rumination, social isolation, and drinking to cope. Psychol Women Q 2016;40(4):564-581 [FREE Full text] [doi: 10.1177/0361684316662603] [Medline: 28138208]
58. Dimova ED, Elliott L, Frankis J, Drabble L, Wiencierz S, Emslie C. Alcohol interventions for LGBTQ+ adults: a systematic review. Drug Alcohol Rev 2022 Jan;41(1):43-53. [doi: 10.1111/dar.13358] [Medline: 34333818]
59. Kidd JD, Paschen-Wolff MM, Mericle AA, Caceres BA, Drabble LA, Hughes TL. A scoping review of alcohol, tobacco, and other drug use treatment interventions for sexual and gender minority populations. J Subst Abuse Treat 2022 Feb;133:108539 [FREE Full text] [doi: 10.1016/j.jsat.2021.108539] [Medline: 34175174]
60. Allen JL, Mowbray O. Sexual orientation, treatment utilization, and barriers for alcohol related problems: findings from a nationally representative sample. Drug Alcohol Depend 2016 Apr 01;161:323-330. [doi: 10.1016/j.drugalcdep.2016.02.025] [Medline: 26936411]
61. Green KE. Barriers and treatment preferences reported by worried drinkers of various sexual orientations. Alcohol Treat Q 2011;29(1):45-63. [doi: 10.1080/07347324.2011.538311]
62. Levak S, Kuerbis AN, Morgenstern J. Drink goal difficulty effect on outcomes in moderation-based alcohol treatment for sexual minority men. J Subst Abuse Treat 2020 May;112:1-9 [FREE Full text] [doi: 10.1016/j.jsat.2020.01.001] [Medline: 32076362]
63. Ratts MJ, Kaloper M, McReady C, Tighe L, Butler SK, Dempsey K, et al. Safe space programs in K-12 schools: creating a visible presence of LGBTQ allies. J LGBT Issues Couns 2013;7(4):387-404. [doi: 10.1080/15538605.2013.839344]
64. Riggle ED, Gonzalez KA, Rostosky SS, Black WW. Cultivating positive LGBTQA identities: an intervention study with college students. J LGBT Issues Couns 2014;8(3):264-281. [doi: 10.1080/15538605.2014.933468]
65. Hatzenbuehler ML. How does sexual minority stigma "get under the skin"? A psychological mediation framework. Psychol Bull 2009 Sep;135(5):707-730 [FREE Full text] [doi: 10.1037/a0016441] [Medline: 19702379]
66. Lick DJ, Durso LE, Johnson KL. Minority stress and physical health among sexual minorities. Perspect Psychol Sci 2013 Sep;8(5):521-548. [doi: 10.1177/1745691613497965] [Medline: 26173210]
67. Sparkman G, Walton GM. Dynamic norms promote sustainable behavior, even if it is counternormative. Psychol Sci 2017 Nov;28(11):1663-1674. [doi: 10.1177/0956797617719950] [Medline: 28961062]
68. Sparkman G, Walton GM. Witnessing change: dynamic norms help resolve diverse barriers to personal change. J Exp Soc Psychol 2019 May;82:238-252. [doi: 10.1016/j.jesp.2019.01.007]

Abbreviations

AUD: alcohol use disorder
CONSORT: Consolidated Standards of Reporting Trials
LBQ: lesbian, bisexual, and queer
PNF: personalized normative feedback
RCT: randomized controlled trial
T1: time point 1
T2: time point 2
T3: time point 3

Edited by T Leung; submitted 11.11.21; peer-reviewed by L Drabble, T Fazzino, H Mehdizadeh; comments to author 28.01.22; revised version received 14.02.22; accepted 21.02.22; published 13.05.22.

Please cite as:
Boyle SC, LaBrie JW, Trager BM, Costine LD
A Gamified Personalized Normative Feedback App to Reduce Drinking Among Sexual Minority Women: Randomized Controlled Trial and Feasibility Study
J Med Internet Res 2022;24(5):e34853
URL: https://www.jmir.org/2022/5/e34853
doi:10.2196/34853
PMID:35559854

©Sarah C Boyle, Joseph W LaBrie, Bradley M Trager, Lauren D Costine. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.05.2022. 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

XSL•FO
RenderX

complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Telephone and Web-Based Delivery of Healthy Eating and Active Living Interventions for Parents of Children Aged 2 to 6 Years: Mixed Methods Process Evaluation of the Time for Healthy Habits Translation Trial

Megan L Hammersley[1,2,3], PhD; Rebecca J Wyse[4,5,6,7], PhD; Rachel A Jones[1,2,8], PhD; Anthony D Okely[1,2,3], EdD; Luke Wolfenden[4,5,6,7], PhD; Simon Eckermann[2,3], PhD; Joe Xu[9], PhD; Amanda Green[9], MPH; Fiona Stacey[4,5,6,7], PhD; Sze Lin Yoong[10], PhD; Jacklyn Jackson[4,5,6,7], PhD; Christine Innes-Hughes[9], MPH; Vincy Li[11], MPhil; Chris Rissel[12], PhD

[1]Early Start, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia
[2]Illawarra Health and Medical Research Institute, Wollongong, Australia
[3]School of Health and Society, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia
[4]School of Medicine and Public Health, Faculty of Health and Medicine, University of Newcastle, Callaghan, Australia
[5]Hunter New England Population Health, Wallsend, Australia
[6]Hunter Medical Research Institute, New Lambton Heights, Australia
[7]Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia
[8]School of Education, Faculty of the Arts, Social Sciences and Humanities, University of Wollongong, Wollongong, Australia
[9]Centre for Population Health, New South Wales Ministry of Health, St Leonards, Sydney, Australia
[10]School of Health Sciences, Swinburne University of Technology, Hawthorn, Melbourne, Australia
[11]Health Consult, Sydney, Australia
[12]College of Medicine and Public Health, Flinders University, Darwin, Australia

Corresponding Author:
Megan L Hammersley, PhD
Early Start
Faculty of the Arts, Social Sciences and Humanities
University of Wollongong
Northfields Ave
Wollongong, 2522
Australia
Phone: 61 242215670
Email: mhammers@uow.edu.au

Abstract

Background: Few translational trials have provided detailed reports of process evaluation results.

Objective: This study reported on findings from a mixed methods process evaluation of a large translational trial comparing 2 remotely delivered healthy eating and active living interventions with an active control, targeting parents of young children.

Methods: Mixed methods process evaluation data were collected as part of a 3-arm, partially randomized preference trial targeting parents of children aged 2 to 6 years from New South Wales, Australia. Recruitment strategies were assessed through the participant baseline questionnaire and a questionnaire completed by the health promotion staff involved in recruitment. Data on participants' intervention preferences were collected at baseline and after the intervention. Intervention acceptability and demographic data were collected via a postintervention questionnaire (approximately 3 months after baseline), which was supplemented by qualitative participant interviews. Implementation data on intervention fidelity and withdrawal were also recorded. Differences in intervention acceptability, fidelity, and withdrawal rates between telephone and web-based interventions and between randomized and nonrandomized participants were analyzed. The significance level was set at $P<.05$ for all tests. The interview content was analyzed, key themes were drawn from participant responses, and findings were described narratively.

XSL•FO
RenderX

Results: Data were collected from 458 participants in the baseline survey and 144 (31.4%) participants in the 3-month postintervention survey. A total of 30 participants completed the qualitative interviews. A total of 6 health promotion staff members participated in the survey on recruitment strategies. Most participants were recruited from Early Childhood Education and Care services. There was a broad reach of the study; however, better take-up rates were observed in regional and rural areas compared with metropolitan areas. Parents with a university education were overrepresented. Most participants preferred the web-based medium of delivery at baseline. There was high acceptability of the web-based and telephone interventions. Participants found the healthy eating content to be the most useful component of the modules (web-based) and calls (telephone). They regarded text (web-based) or verbal (telephone) information as the most useful component. A high proportion of participants completed the telephone intervention compared with the web-based intervention; however, more participants actively withdrew from the telephone intervention.

Conclusions: This is one of the first studies to comprehensively report on process evaluation data from a translation trial, which demonstrated high acceptability of all interventions but a strong participant preference for the web-based intervention. This detailed process evaluation is critical to inform further implementation and be considered alongside the effectiveness outcomes.

(J Med Internet Res 2022;24(5):e35771) doi:10.2196/35771

KEYWORDS

dietary intake; physical activity; screen time; sleep; movement behaviors; online; internet; telephone; mobile phone

Introduction

The dietary habits and movement behaviors (physical activity, screen time, and sleep) of Australian children are well below the current recommendations and have deteriorated over time [1]. Parents are fundamental to establishing healthy behaviors during early childhood [2]. However, there are several barriers that can impede the involvement of parents in healthy lifestyle interventions for their children [3,4]. Remotely delivered interventions, such as telephone or web-based programs, have the potential to overcome these barriers, allow access regardless of location, and provide greater flexibility compared with face-to-face interventions. The Healthy Habits and Time2bHealthy remotely delivered parent-focused interventions have demonstrated efficacy in randomized controlled trials (RCTs). The *Healthy Habits* 4-week telephone-based intervention for parents of children aged 3 to 5 years showed a significant improvement in children's fruit and vegetable intake [5]. The T*ime2bHealthy* 11-week web-based intervention demonstrated significant improvement in children's discretionary food intake, parental nutrition self-efficacy, and child feeding practices [6]. These interventions were conducted under highly controlled conditions; however, more recently, both have been tested in a large translation trial (known as *Time for Healthy Habits*) to determine their effectiveness in a real-world context, with the potential for widespread implementation [7]. The *Time for Healthy Habits* study [8] investigated the effectiveness of the *Healthy Habits Plus* (enhanced *Healthy Habits*) telephone intervention and the *Time2bHealthy* web-based intervention against an active control group (receiving written materials). The protocol [8] and main outcomes of the *Time for Healthy Habits* translation study have been published elsewhere. Briefly, the study found that although there was no statistically significant difference between groups over time in relation to the primary outcome (children's fruit and vegetable intake), there was a significant improvement over time among randomized participants receiving the telephone intervention for non–core food intake (secondary outcome) compared with participants receiving the control (written materials) [9]. There is a need to evaluate process data to further explore and explain these results so that any future decisions related to the potential scale-up of these interventions are fully informed. Process evaluations are critical to providing a comprehensive assessment of interventions alongside effectiveness testing, helping to determine how interventions work, whom they work for, how outcomes can be explained, and how interventions can be improved in the future, which are important considerations for policy and practice [10]. To date, a very limited number of studies have conducted process evaluations of children's healthy eating and active living translation trials [11-13]. This study aimed to determine intervention acceptability, optimal recruitment strategies, participant intervention preference (ex ante and ex post), intervention fidelity, withdrawal rates, and participant representativeness concerning the target population.

Methods

Study Overview

This was a process evaluation of the *Time for Healthy Habits* study and comprised participant data from the main trial (collected at baseline and 3 months after the intervention) and data from 30 qualitative interviews across all intervention arms, conducted 1 to 10 months after the intervention. It also comprised data collected from the participant recruitment staff in the local health districts (LHDs) where the main trial was conducted. A detailed description of the protocol for the main effectiveness trial has been previously published [8]. Briefly, parents of children aged 2 to 6 years from New South Wales (NSW), Australia, were recruited. Parents were eligible if their children lived with them for at least 4 days per week on average and they spoke sufficient English to participate. The trial design was a 3-arm, partially randomized preference trial. Participants were initially provided with the option to choose their preferred delivery method (telephone, web-based, or written material) or to be randomized. This allowed us to establish the participants' ex ante intervention preferences. The design was also thought to have higher initial participant acceptability than a traditional RCT, as participants may have been more willing to take part and complete the intervention if they knew that they were able

XSL•FO
RenderX

to choose which intervention they received [14-19]. However, to ensure that sufficient participants were enrolled in the randomized arm of the study to establish intervention effectiveness via robust analysis, a stopping rule was applied to limit the number of participants who could choose their preferred intervention. After the application of the stopping rule, all participants were randomized in a 1:1:1 ratio. The 3 arms of the study included *Healthy Habits Plus* (a telephone intervention), *Time2bHealthy* (a web-based intervention), and an active control (written education materials). The specific features of the interventions are outlined in Table 1.

Table 1. Time for Healthy Habits intervention components.

Components	Time2bHealthy (web-based)	Healthy habits plus (telephone)	Active control (written materials)
Format	• Web-based web application comprising 6 modules (1 per fortnight) over 3 months with email reminders	• Six 20- to 30-minute fortnightly telephone calls over 3 months	2 factsheets fortnightly (10 in total) and 1 summary booklet over a period of 3 months
Content	• Text, videos, practical activities, and quizzes • Optional closed Facebook group	• Verbal information • Guidebook containing additional information and resources • Pad of meal planner templates	Text information and images
Behavior change strategies	• Barrier identification, goal setting, and self-monitoring	• Barrier identification, goal setting, and self-monitoring	N/A[a]
Topics	• Healthy eating, physical activity, screen time, and sleep	• Healthy eating, physical activity, screen time, and sleep	Healthy eating, physical activity, screen time, and sleep

[a]N/A: not applicable.

Ethics Approval

Ethics approval for the quantitative and qualitative aspects of this study was granted by the South Western Sydney LHD Human Research Ethics Committee (HE18/300), and site-specific approval was obtained from the human research ethics committees of the 5 LHDs involved in the study [8]. Acceptance was provided by the University of Newcastle Human Research Ethics Committee (H-2019-0188) and the University of Wollongong Human Research Ethics Committee (HE2019/207).

Process Evaluation Data Collection and Measures

This mixed methods process evaluation reported data from the sources detailed in the following sections.

Preference and Demographic Characteristics

Baseline Questionnaire

Ex ante preferences were collected from all trial participants during the baseline interview (via telephone). Before the implementation of the stopping rule, participants were asked, "Do you have a strong preference for the way in which you receive healthy lifestyle advice or support about your child?" If they responded *yes*, they were then asked, "Would you prefer to receive healthy lifestyle advice or support via written information, telephone, or online" (with the order in which the interventions were stated to be randomized). After the implementation of the stopping rule, participants were still asked what their preferences would have been, although all participants were randomized from this point. Basic demographic data were also collected.

3-Month Postbaseline Questionnaire

Ex post preference was ascertained from participants by the following question: "Having completed the program, would you have preferred for the information to be delivered in another way?" If they responded *yes*, they were asked, "In which format would you have preferred to receive the advice?" (response options included *online program*, *telephone counseling*, *educational materials*, *smartphone app*, *face-to-face*, *Skype*, *other*, and *do not know*).

Recruitment: Health Promotion Staff Surveys

The LHD staff (recruitment officers or other health promotion staff who were involved in the recruitment of parents to the study) completed a web-based questionnaire comprising 10 questions. The questions focused on recruitment strategies; recruitment challenges; and recommendations for future recruitment, including additional support. These questions are provided in Multimedia Appendix 1. In addition, the baseline participant questionnaire included a question on where they heard about the study.

Intervention Acceptability

3-Month-Postbaseline Questionnaire

The postintervention (3 months after baseline) questionnaire included up to 27 process evaluation questions (depending on the intervention) and was completed over the telephone (for *Healthy Habits Plus* participants), on a web-based questionnaire (for *Time2bHealthy* and control group participants). A complete list of process evaluation questions can be found in Multimedia Appendix 2. The process evaluation questions were similar to those used previously in the process evaluation of the *Time2bHealthy* (web-based) RCT [6] and measured user acceptance of the content and modality of each intervention. Specifically, the participants were asked 5 questions about whether the intervention content was interesting, easy to understand, relevant to their family, worthwhile, and had information that they could act on. These questions used a Likert

XSL•FO
RenderX

scale, with semantic anchors ranging from 1 (strongly disagree) to 5 (strongly agree). Each of these question responses was summed to attain an overall user acceptance score.

Participants were also asked about the appropriateness of the length and number of calls, web-based modules, or written resources. Furthermore, regarding the telephone and web-based interventions, they were asked to identify 1 intervention aspect that they found most useful (eg, for the telephone intervention: the guidebook, information provided verbally by the interviewer, goal setting, homework activities, and the meal planner templates; for the web-based intervention: information provided in text, videos, goal setting, and activities). Participants were also asked to identify the 1 call or module that they found most useful.

Participant Qualitative Interviews

In addition to the abovementioned questions asked in the 3-month postbaseline follow-up, a sample of participants from each of the 3 interventions was invited to participate in an additional telephone interview to further explore participants' experiences. Participants were selected from a list of all those who participated in the interventions by March 2020, with the aim of interviewing 10 participants per intervention and the intention of capturing a targeted selection from metropolitan, rural, and regional areas (target of 16 metropolitan, 7 regional, and 7 rural participants); a combination of participants who had partially and fully completed the interventions; and a mix of participants from the randomized and preference arms of the study. Interviews were conducted by a research consulting company (Research Forum Consulting), which emailed participants, provided details about the interviews along with a participant information sheet, and informed them that they might receive a phone call to invite them to participate. When participants were phoned, they were provided with information about the interviews and asked to participate. Consent to proceed with the interview was obtained verbally, and participants provided consent for the interview to be audio recorded (optional). Questions were designed to capture participants' overall and intervention-specific experiences (for telephone and web-based intervention participants only). Participants were asked about their initial expectations, intervention content, what they found most and least useful, length of the interventions, ease of completion, and engagement. A copy of the interview questions can be found in Multimedia Appendix 3. A total of 30 participants were interviewed (10 from each intervention group), which was anticipated to represent participants' breadth of experience.

Fidelity and Withdrawal Rates: Intervention Implementation Data

Although the interventions were designed to be completed within 12 weeks, additional time was allocated (up to 20 weeks in total) to allow participants to complete the interventions by extending access to the web-based intervention and continuing to contact telephone participants to complete the intervention calls that had not yet been completed. Data were collected on the withdrawal of participants from the study, including whether the withdrawal was active (where the participant explicitly asked to be withdrawn) or passive (where the participant did not complete the intervention but did not ask to be withdrawn). We also determined the proportion of participants who completed each phase of the intervention; that is, the number of calls or modules completed.

Data Analyses

Key themes were drawn from the responses to the health promotion staff survey questions, and findings were described narratively in relation to recruitment avenues used, recruitment barriers, and strategies that were most and least successful for recruitment. Participant responses to the 3-month postbaseline Likert scale questions on user acceptance were considered singularly and summed to produce a score from 5 to 25. Medians and IQR were determined. Kruskal-Wallis tests were used to assess differences in Likert scale responses to questions between all study groups and between randomized and nonrandomized participants. Participants' qualitative interviews were audio recorded for all participants who provided consent (28/30, 93%) and then transcribed verbatim and deidentified. Detailed notes were obtained for those who did not consent to be recorded (2/30, 7%). The interview content was analyzed, and key themes were drawn from the participant responses. These findings were then described narratively in relation to the specific question domains, which were triangulated with the quantitative participant questionnaire data relating to the participant acceptability of the interventions. The number and percentage of preferences, randomized and total participants by study arm (telephone, web-based, and active control written materials) completing each module or call, and the number and percentage of active and total withdrawals (active and passive withdrawals) were calculated. Chi-square tests were used to assess differences in active and total withdrawal rates between interventions and differences between randomized and nonrandomized participants. Mann-Whiney *U* tests were used to assess differences in the number of calls or modules completed between the telephone and web-based study groups and differences between randomized and nonrandomized participants. Completion numbers and percentages (both completion at any time point and completion within 20 weeks) were calculated by the intervention group according to whether participants were randomized or nonrandomized. Chi-square tests were used to assess differences in completion of the intervention between the telephone and web-based intervention groups and differences between randomized and nonrandomized participants. In the first instance, these tests were based on the completion of modules, calls, and interventions within any time frame. They were then repeated based on the completion of the modules or calls within a 20-week time frame. The significance level was set at $P<.05$ for all tests. All analyses were conducted using SPSS Statistics for Windows (version 25.0; IBM Corp).

Results

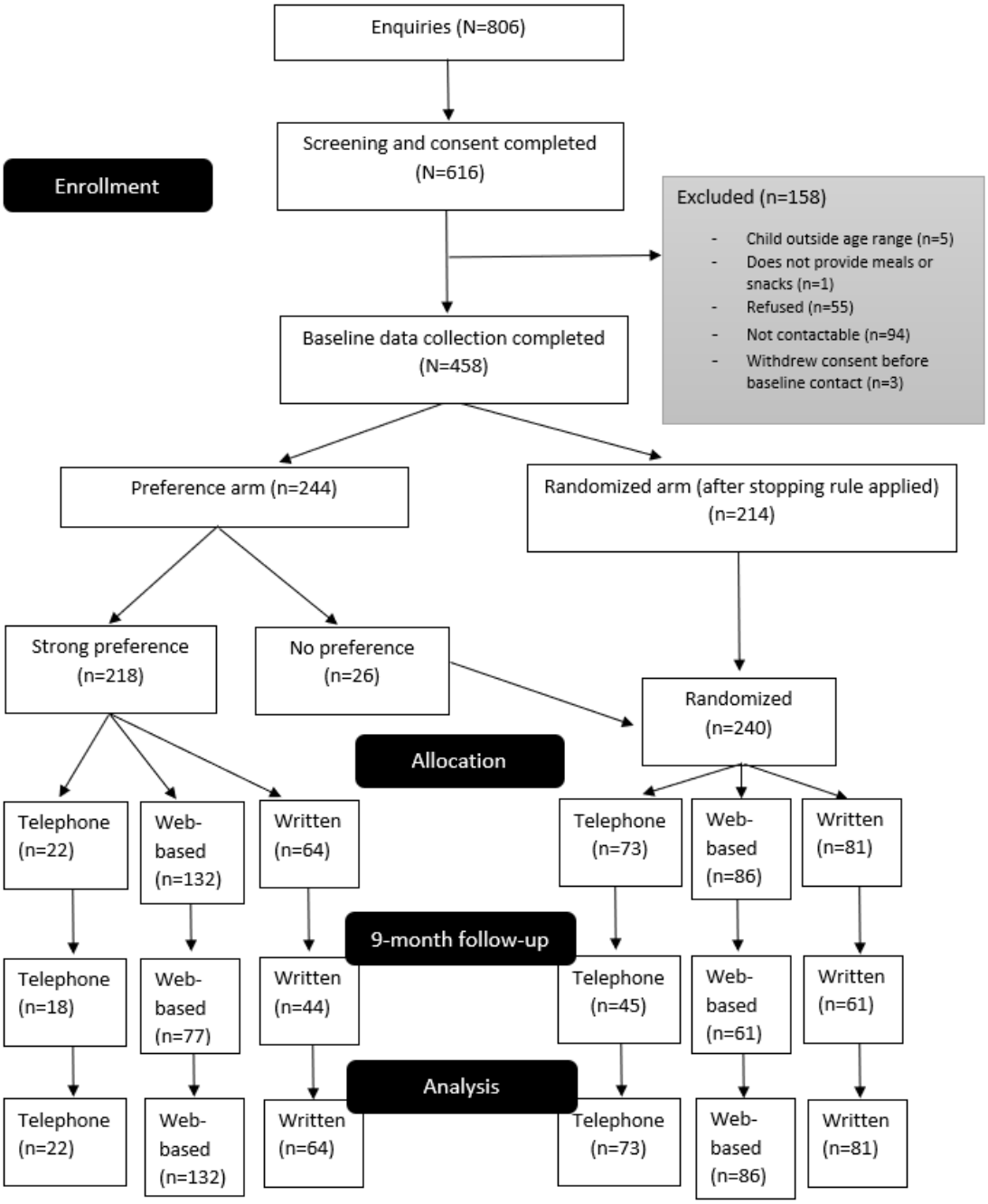
Overview

Data collection was completed by 458 participants at baseline and 144 (31.4%) at postintervention (3 months after baseline), as shown in Figure 1. Of the 79 invited participants, 30 (38%) completed qualitative interviews, 10 (13%) from each intervention (including the active control group). Interviews for

XSL•FO
RenderX

the qualitative study were conducted between March and April 2020, which was between 1 and 8 months after participants had completed one of the interventions (or the active control). Approximately 80% (24/30) were from metropolitan areas, and 20% (6/30) were from regional and rural areas (there was a higher proportion of metropolitan participants than the target). There was an even split of randomized to preference participants for the telephone (5:5) and active control (written materials; 4:6) participants; however, only one of the participants from the list of those who had completed the web-based intervention was randomized; thus, 30% (9/30) of participants interviewed from this intervention group were preference participants. Although attempts were made to include participants who had only partially completed the interventions, all participants who were interviewed had completed the interventions. All 5 LHDs targeted for recruitment were represented in the 6 responses to the health promotion staff survey.

Figure 1. Time for healthy habits process evaluation CONSORT (Consolidated Standards of Reporting Trials) diagram.



Recruitment

Participants were recruited for the main trial using established networks such as health promotion staff visits to Early Childhood Education and Care (ECEC) services; Child and Family Health nurses; and several other avenues, including playgroups and libraries, social media, media releases, the Playgroup Australia newsletter, the University of Wollongong Discovery Space (children's museum) newsletter, and through health professionals. Most LHD-based recruitment works were conducted directly by recruitment staff employed within each LHD for the specific purpose of assisting with recruitment to

XSL•FO
RenderX

this trial, with other recruitment activities (such as social media, newsletters, and bulk emails) conducted by a central project coordinator.

Participants reported finding out about the study from ECEC services or educators (188/458, 41%), Facebook or social media (87/458, 19%), and libraries (23/458, 5%). These findings were reiterated by the LHD recruitment staff, who indicated that ECEC services were the most successful channel, with some LHDs reporting that this was particularly effective when attending face to face, where there was an opportunity to talk to parents about the study directly. They also reported that they found library groups, playgroups, large events, and preschool sporting activities as efficient recruitment sites (likely because of being able to speak face to face with parents), and some (but not all) reported that mass mailouts to ECEC services were effective in recruiting parents.

The LHD health promotion staff reported via the survey that a facilitator of these successful recruitment channels was the ability to answer questions from parents face to face. They felt that parents were more likely to enroll in the study when there was someone present face to face, as many parents reported that they had seen a flyer about the study before enrolling:

> *Face-to-face conversation with parents, being able to explain the program to them in detail. Providing parents with the opportunity to ask questions and seek more info before signing up is important.*

The LHD staff reported that the least effective recruitment channels were emails, posters, flyers, media releases, and large events that were not targeted to the age group. Although some staff members found mass mailouts to ECEC services effective, others did not. Some stated that mailouts were more successful when accompanied by a follow-up telephone call for the ECEC service:

> *Email alone—rarely received any form of contact/enquiry. Slightly better if emails were followed up with a phone call.*

Demographics

Considering the broad representativeness of the trial, there was a substantial representation of NSW-target LHDs in regional and rural areas; however, there was an underrepresentation of LHDs in metropolitan areas. The breakdown of participant proportions across the target LHDs is displayed in Table 2. Parent participants in the study were more likely to be female (441/458, 96.3%) compared with the general NSW population (50.7%). The mean age of participants (36.13, SD 4.92 years) was similar but perhaps slightly older than the general NSW parent population, given the median age of NSW first-time mothers and fathers (30.7 and 33.1 years, respectively) and that some parents already had older children. The mean age of child participants was 3.37 years (SD 1.16). A smaller percentage of participants spoke a language other than English at home (81/458, 17.7%) compared with the general NSW population (27%). The proportion of Aboriginal and Torres Strait Islander individuals (15/448, 3.3%) was similar to that of the NSW population (2.9%). There was a much higher proportion of university-qualified participants (322/458, 70.3%) than in the NSW population (23.4%). Over three-fourth of the participants had a household income higher than the NSW median household income [20].

Table 2. Number and proportion of participants recruited across target LHDs[a] compared with drawing area (N=380).

LHD	Geographic area	Children in drawing area and proportion of total target drawing area[b], n (%)	Participants recruited[c], n (%)
Illawarra Shoalhaven	Regional and rural	49,791 (15.73)	121 (31.8)
Murrumbidgee	Regional and rural	23,133 (7.31)	42 (11.1)
Southern NSW[d]	Regional and rural	24,483 (7.73)	43 (11.3)
Hunter New England	Regional and rural	118,306 (37.37)	83 (21.8)
South Eastern Sydney	Metropolitan	100,826 (31.85)	91 (23.9)

[a]LHD: local health district.

[b]Number of children aged 0 to 9 years in each LHD (statistics on children aged 2 to 6 years unavailable; Center for Epidemiology and Evidence. HealthStats NSW: Population by Local Health District. 2019).

[c]Remaining participants (n=78) were recruited from areas of NSW that were not specifically targeted for recruitment.

[d]NSW: New South Wales.

Intervention Preference

Ex Ante

At baseline, all participants were asked whether they had a strong preference for how they received health information. Of the 458 participants, 393 (85.8%) stated that they had strong preferences. When asked which delivery medium they preferred, 59.3% (233/393) stated *online*, 28.5% (112/393) stated *written materials*, 11.9% (47/393) stated *telephone calls*, and 0.3% (1/393) stated that they did not know.

Ex Post

When asked in the postintervention process evaluation if they would have preferred to receive the intervention in another way, 30.5% (44/144) of the respondents stated that they would have. This included 44% (35/80) of the randomized arm and 14% (9/64) of the preference arm. Further details on the ex post intervention preferences of participants by intervention group and study arm are shown in Table 3. The most commonly stated alternative delivery media preferences were digital delivery

XSL•FO
RenderX

mediums such as smartphone apps (10/144, 6.9%) or the web (9/144, 6.3%).

Table 3. Ex post intervention preferences of participants by intervention and study arm (N=144).

Intervention	Total participants preferring alternative delivery method		Randomized arm			Preference arm		
	Total sample	Participants, n (%)	Participants preferring alternative delivery method		Preferred method stated[a]	Participants preferring alternative delivery method		Preferred method stated[a]
			Total sample	Participants, n (%)		Total sample	Participants, n (%)	
Telephone	48	13 (27)	35	11 (31)	• Web-based (n=9) • Smartphone app (n=1) • Educational materials (n=1)	13	2 (15)	• Phone and web-based (n=2)
Web-based	57	9 (16)	21	6 (29)	• Smartphone app (n=2) • Podcast (n=1) • Telephone (n=1)	36	3 (8)	• Smartphone app (n=3) • Educational materials (n=1)
Active control (written materials)	39	22 (56)	24	18 (75)	• Telephone (n=3) • Face-to-face (n=1) • Smartphone app (n=2)	15	4 (27)	• Face-to-face (n=2) • Smartphone app (n=2) • Skype (n=1)
Total	144	44 (31)	80	35 (44)	N/A[b]	64	9 (14)	N/A

[a]Not all participants who preferred an alternative delivery method stated what their preference was, and some participants provided >1 option; hence, some numbers do not add up to the total.

[b]N/A: not applicable.

Intervention Acceptability

The 3-month postintervention process evaluation found that there was a high level of acceptability for all the interventions, with the median overall score for participants being 22.0 (IQR 5.0) out of a possible high score of 25 (Table 4). The highest median overall score was obtained for the telephone intervention (23.0, IQR 4.0), with the web-based (22.0, IQR 5.0) and active control (written materials; 22.0, IQR 4.0) interventions being similar. There was a significantly higher score for the telephone intervention than that of the active control (written materials) regarding overall acceptability (H_1=8.258; *P*=.004), the intervention being regarded as interesting (H_1=9.176; *P*=.002), worthwhile (H_1=8.878; *P*=.003), and having information that participants could act on (H_1=10.044; *P*=.002). There was also a significantly higher score for the web-based intervention compared with the active control (written materials) about being regarded as worthwhile (H_1=6.299; *P*=.01) and having information that participants could act on (H_1=5.548; *P*=.02).

Participants who completed the in-depth telephone interviews commented that the modules or calls were easy to follow. When asked to rate the interventions on a scale of 1 to 10 (1=very easy to 10=extremely hard), they rated the telephone (mean 3.5, SD 2.2) and web-based (mean 3.1, SD 2.4) interventions similarly. Challenges experienced by participants were usually not related to the interventions per se but rather to the implementation of changes with their children.

XSL•FO
RenderX

Table 4. Participant feedback on intervention acceptability from the 3-month postintervention questionnaire (N=144).

	Web-based, median (IQR)			Telephone, median (IQR)			Active control (written materials), median (IQR)			All participants, median (IQR)		
	Randomized (n=21)	Preference (n=15)	All web-based (n=38)	Randomized (n=35)	Preference (n=13)	All telephone (n=48)	Randomized (n=15)	Preference n=23)	All written (n=38)	Randomized (n=79)	Preference (n=64)	Total participants (n=143)
Program was interesting[a]	4.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0[b] (1.0)	4.0 (0.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	5.0 (1.0)	4.0 (1.0)
Program was easy to understand[a]	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (0.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (1.0)	5.0 (2.0)
Program was relevant to family[a]	5.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	5.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)
Program was worthwhile[a]	5.0 (1.0)	5.0 (1.0)	5.0[b] (1.0)	5.0 (1.0)	5.0 (1.0)	5.0[b] (1.0)	4.0 (1.0)	4.0 (2.0)	4.0 (1.0)	4.0 (1.0)	5.0 (1.0)	4.0 (1.0)
Could act on information[a]	5.0 (1.0)	4.0 (1.0)	4.5[c] (1.0)	4.0 (1.0)	5.0 (1.0)	4.5[b] (1.0)	4.0 (0.0)	4.0 (1.0)	4.0 (0.0)	4.0 (1.0)	4.0 (1.0)	4.0 (1.0)
Overall score[c]	24.0 (5.0)	22.0 (5.0)	22.0 (5.0)	23.0 (4.0)	23.0 (4.0)	23.0[b] (4.0)	20.0 (3.0)	22.0 (6.0)	22.0 (4.0)	22.0 (5.0)	22.0 (5.0)	22.0 (5.0)

[a]Likert scale of 1 to 5 for each question.

[b]Significant difference between intervention and control (P<.05).

[c]Overall score was the sum of all scores (possible range 5-25).

Web-Based Intervention

Overview

Almost all web-based participants in the postintervention process evaluation suggested that the number of modules (ie, n=6) was appropriate (53/56, 95%) and that the length of each module was appropriate (55/57, 96%):

> *Yeah, I think it was just right. There was a couple of busy weeks for myself so I may have done part of the module each day over the week and then I still had plenty of time to implement my goals.* [Participant 2, mother of boy aged 4 years]

The web-based intervention comprised the following components: written information, interactive activities, goal setting, videos and quizzes, and an optional Facebook group. Participants in the postintervention process evaluation suggested that the most useful intervention components were written information (23/57, 40%), interactive activities (eg, planner, recipe modification, and label reading; 15/57, 26%), and goal-setting components (13/57, 23%). In contrast, the 10 qualitative interviews for the web-based intervention suggested that goal setting, videos, and quizzes were the most useful components:

> *My favourite part was the goal setting. It made you reflect, or it was new information to you and actually saying what you're going to do. It did create a goal for the next couple of weeks and held me accountable to that.* [Participant 2, mother of boy aged 4 years]

> *I guess the content. It kind of gave you a plan. The quizzes I suppose I could redo them so I could refresh and put them in place. I could go back and try this. It was kind of like a tool you could go back to.* [Participant 20, mother of boy aged 5 years]

Of the web-based intervention content (which focused on healthy snacks, healthy meals, physical activity, screen time, and sleep), the highest proportion of participants perceived healthy eating or healthy snacks content as the most useful (27/57, 47%, and 13/57, 23%, respectively) in the postintervention process evaluation. This was supported by the findings from the qualitative part of the study, with most participants stating that healthy eating and physical activity modules were the most useful. The least useful modules were considered to be sleep and screen time as participants thought they had already established good practices in these areas:

> *Before the program I was really struggling to come up with ideas of healthy snacks. Doing it gave me some more ideas, taught me how to read labels and work out what was healthy and what wasn't healthy. It also reinforced that most of what I was doing was right but giving me a few extra tricks, I guess.* [Participant 11, mother of girl aged 4 years]

XSL•FO
RenderX

> *I think the physical activity...Because they had some ideas, things I could play with the children in the yard. I really like this they do play in the yard but just in case they get bored I have some things I can do with them. I found that helpful.* [Participant 13, mother of girl aged 4 years]

Telephone Intervention

Most participants in the postintervention process evaluation suggested that the number of calls in the telephone intervention was appropriate (40/48, 83%) and that the length of the calls was appropriate (44/48, 92%):

> *Yeah, I mean they didn't drag on. They didn't make any points unnecessarily or whatever.* [Participant 17, mother of boy aged 2.5 years]

Participants in the postintervention process evaluation reported that the most useful intervention components of the telephone intervention were the verbal information (20/48, 42%), guidebook (15/48, 31%), and goal setting (10/48, 21%). Regarding the useful content, most participants (38/48, 79%) reported that healthy eating was the most useful. These data are supported by the data from the qualitative part of the study with participants, suggesting that the guidebook and goal setting were the most useful intervention components, and the healthy eating content was the most useful. The least useful content was related to screen time and sleep:

> *I love the guidebook. I thought it was great. Just having that reference, I would look through it before our phone call. I could follow on when having the phone call.*
>
> *I think the accountability side of it. You would pick the goals and then have someone call and follow up and say how are you going with that. That made you think if you hadn't been focussing on it you thought yeah I should be doing more in terms of working towards that goal.* [Participant 17, mother of boy aged 2.5 years]

Participants stated that they benefited from knowledge regarding the amount of physical activity required, goal setting, implementing changes as a family, encouragement of family meals, healthy eating tips, practical advice, and support for implementing changes:

> *The other really helpful thing was thinking about how to encourage good eating behaviours like mealtime behaviours, sitting with the family. That's something we changed as well. We used to make my son eat separately. Now from time to time when we can we sit and eat as a family to model the good eating behaviours.* [Participant 17, mother of boy aged 2.5 years]
>
> *In terms of that yeah. Getting her out on a bike getting her out walking. Really promoting more active play outside because it's not what she would normally tend to show interest towards. That I think has been a lifestyle change for us; it's something we've implemented and stuck to.* [Participant 7, mother of girl aged 4 years]

Active Control (Written Materials)

Approximately three-fourth of the participants in the postevaluation process evaluation suggested that the active control (written materials) components were appropriate in terms of the number of resources and amount of information included (20/27, 74%, and 19/27, 70%, respectively). This was supported by qualitative interview data:

> *Yeah, I think it was great. Well I think with the exercise bit it was quite good to see what's considered exercise as well. I made changes. I've used the tips for the lunchboxes I guess so I yeah I did use some of these ideas.* [Participant 3, mother of girl aged 5 years]
>
> *I feel just increasing vegetable and fruit intake and making meals a bit more fun. I think that's probably the main thing that we've taken from it. Also, enjoying outdoor activities.* [Participant 14, mother of boy aged 3 years]

Fidelity

Significantly more participants in the telephone arm completed the intervention than those in the web-based arm (47/95, 50%, vs 57/218, 26.1%, respectively; $P<.001$). When considering participants who completed the intervention within a 20-week timeframe, there was no significant difference between the 2 groups, with 33% (31/95) of participants completing the telephone intervention and 25.7% (56/218) of participants completing the web-based intervention within 20 weeks. There was no significant difference in intervention completion between the randomized and preference groups. Within the web-based intervention, out of 218 participants, 105 (48.2%) joined the optional Facebook group.

Withdrawal

Although there was a greater proportion of total withdrawals (including active and passive withdrawal) in the web-based versus telephone group (161/218, 73.9%, vs 47/95, 50%; $P<.001$), there was a significantly higher proportion of participants in the telephone intervention group who actively withdrew from the intervention than those in the web-based intervention group (19/95, 20%, vs 4/218, 1.8%; $P<.001$). There was no significant difference in withdrawal rates between the randomized and preference participants.

Discussion

Principal Findings

A comprehensive process evaluation of the *Time for Healthy Habits* translation study of 2 remotely delivered healthy eating and active living interventions and an active control for parents of children aged 2 to 6 years was conducted in this mixed methods study. There was a broad reach of the study across metropolitan, regional, and rural areas of NSW, Australia; however, there were better take-up rates in some areas of the state than in others, with higher participation rates in regional and rural areas than in metropolitan areas. The recruitment effort through the LHDs was substantial. Engagement with existing health promotion staff was crucial for recruitment, as they had

XSL•FO
RenderX

established networks with ECEC services to facilitate recruitment. Recruitment was also assisted by having specifically appointed recruitment staff within the target LHDs, who also concentrated largely on ECEC services for recruitment, resulting in 41% (188/458) of the participants being recruited through this channel. This appeared to be particularly useful when the staff attended face-to-face sessions. Social media was another important avenue of recruitment, where 19% (87/458) of the participants found out about the study. After the preference arm was closed, it was perceived that parents were reluctant to be involved as they did not want to be randomized; however, it was difficult to determine whether this was the result of recruitment saturation over time. There was also limited time and capacity for the health promotion staff to be involved in recruitment. Without additional staff resources available in this trial, it is unlikely that recruitment rates would have reached the same level. This is a common issue for translation trials, where dedicated staff are needed to recruit to a program or service, and it is difficult to obtain a sense of true real-world uptake of such interventions. It is possible that future implementation of interventions in LHDs may result in lower uptake rates; however, some parents may also be more inclined to participate in a program if they are not in need to sign up to a research trial.

Although a larger proportion of participants initially preferred a web-based delivery mechanism, the acceptability of both the web-based and telephone interventions was significantly higher than that of the active control (written materials) regarding being worthwhile and containing information that participants could act on. The telephone intervention also demonstrated a significant difference in acceptability compared with the control for the overall acceptability score, taking into account a wider range of acceptability factors. The ease of following the interventions was similar; however, this rating was slightly better for participants who received the web-based intervention than for those who received the telephone intervention. Regarding the components of the intervention, the web-based participants stated that the text information and goal setting were the most useful, whereas the telephone participants felt that the verbal information and guidebook were the most useful. The usefulness of the text and guidebook information may be influenced by the sample being highly educated, and this may not be generalizable to lower socioeconomic populations where literacy levels are known to be lower [21]. Less than half of the participants receiving the web-based intervention joined the optional Facebook group, and similar to previous studies that have used Facebook as a component of an intervention [22], engagement in the discussion was quite low.

Regarding content, the healthy eating aspects were the most useful across all interventions, with sleep and screen time being regarded as the least useful as they felt these were areas in which their children were already doing well, which is fairly consistent with the current evidence concerning these behaviors in that more young children are meeting the movement behavior guidelines than the dietary guidelines, with vegetable intake, in particular, being very low [1]. Past research also indicates that many parents perceive that their young children are naturally active [23]. However, there is still a great need for improvement in relation to physical activity, as the proportion of children meeting the guidelines drops from 75% at the age of 2 to 3 years to 43% at the age of 4 to 8 years [1]. A small number of participants indicated that they regarded limiting screen time as important. The participant interviews were largely conducted before the COVID-19 pandemic restrictions in 2020. Studies investigating the health behavior habits of children during lockdown periods have indicated that screen time has become a considerable concern that could lead to long-term increased use [24]. Therefore, it is possible that parental concern regarding screen time may have increased since this time and could be in greater need of focus in the future.

For participants in both interventions, engagement declined over time, particularly for those receiving web-based intervention. It is important to address reduced engagement and participation levels, as implementation levels have been demonstrated to have an impact on study outcomes [25]. As suggested in previous studies, offering participants flexibility and choice of delivery medium may assist in uptake and engagement in interventions [26], and it may be worthwhile to consider alternate delivery options in future studies. Most participants who identified an alternate delivery means specifically identified a smartphone app; however, it should be noted that previous research has indicated that apps can also have high attrition rates [27]. Some participants commented that a combination of delivery mechanisms such as telephone and web-based or telephone and smartphone apps would be preferable. Many of the participants who completed the interviews stated that they would also like to receive ongoing support to help embed knowledge and sustain their practices. It is likely that different mediums are a matter of individual preference, and providing multiple options to access interventions may be beneficial when scaling to a population level; however, the practicalities and costs of offering multiple mediums would need to be considered carefully. The only significant outcome of this study was in relation to children's dietary intake of noncore foods. This may be because the healthy eating modules or calls were completed first and by a higher proportion of participants. By the same token, the reason for no significant outcomes for physical activity, screen time, or sleep may be as these topics were covered later, and as the interventions needed to be completed sequentially and engagement dropped off over time, fewer participants completed these calls or modules. Completing the calls and modules sequentially may not be suitable for all participants, and it may be preferable to allow participants to choose their main topics of interest or concern and complete them first.

There was a high withdrawal rate for the interventions, particularly the web-based intervention. Although significantly more telephone participants actively withdrew from the intervention, this may have been because of the nature of the intervention, whereby regular telephone contact was required, and thus, participants needed to actively withdraw if they did not want to receive further phone calls. Previous translation trials have reported challenges with withdrawal and retention [12,28,29]. Several parents cited a lack of time to participate, a common barrier highlighted in previous studies involving parents [26]. There was difficulty contacting some participants

XSL•FO
RenderX

receiving the telephone intervention, and although several attempts were made, some participants could not be contacted.

Parents with a high level of education were overrepresented in the study, a challenge that has been described in other similar studies [30]. A previous study, Healthy Habits, Happy Homes Scotland, effectively engaged lower socioeconomic families, with 65% of the participants living in the most deprived areas. This was achieved using participatory and inclusive strategies, making strong connections with parents and supporting organizations, and coproducing the intervention [11]. It is important that interventions are designed to be acceptable and accessible to lower socioeconomic families so that they are adequately represented in studies, or there is a danger of the gap in health outcomes widening [31]. Other translation studies on older children have effectively reached lower socioeconomic families, with most of the participants recruited through schools or self-referrals [12,13]. However, families from low socioeconomic backgrounds can be less likely to complete these interventions [13]. Given the successful recruitment of participants from ECEC services in this study, focusing on ECEC services in specific postcodes with a high level of social disadvantage may be an effective strategy for engaging families from lower socioeconomic backgrounds in the future. There was also a much higher representation of mothers than fathers. Often, mothers are the primary caregivers of children at this age, which is unsurprising. Previous research has also found that one of the barriers to fathers participating in research is the relative lack of time and availability relative to mothers [32]. In addition, an inclusion criterion was that the child needed to live with the parent for at least 4 days per week to have the opportunity to influence child behaviors, which may have prevented the participation of some fathers with joint custody arrangements. There is evidence that fathers can have a profound influence on the dietary intake and physical activity habits of their children; therefore, it is important that future studies consider the engagement of fathers and ensure that interventions are relevant and accessible to them [33,34].

Strengths and Limitations

A strength of this study was the use of a comprehensive combination of quantitative and qualitative methods to evaluate the process of delivering the interventions, which attempted to obtain diverse representations across the LHDs involved. The qualitative interviews were conducted by a separate research organization; therefore, participants may have been more likely to provide more honest responses to the questions asked. The evaluation was limited by the modest proportion of participants (144/458, 31.4%) who completed the process evaluation questionnaire after the intervention. The sample may have been biased toward those who completed the intervention, and the views of participants who did not complete the interventions may differ from those who completed the interventions. Although all participants were asked to complete the questionnaire around 3 months after baseline, regardless of whether they had completed the interventions, most participants who completed the questionnaire had finished the interventions. Similarly, despite efforts to engage participants who had not completed the interventions in the qualitative interviews, all participants who completed the interviews had completed the interventions; therefore, it was difficult to ascertain the specific reasons for the noncompletion of the interventions. Although the interviews were conducted by a separate research organization, it is possible that participants may not have given their honest opinions. Parents with a university education were overrepresented in the study; therefore, these process evaluation results may not be representative of the general population. Finally, this study was conducted during a period that encompassed the height of the 2020 COVID-19 pandemic restrictions in NSW from mid-March to late May 2020, when most school children were at home learning remotely, ECEC services were encouraging children to stay at home, and many parents were working from home. Recruitment of participants was likely affected by these restrictions, and anecdotal reports indicate that some parents found the additional time pressures during this period difficult, and completion of the interventions and the implementation of behavior changes may have been affected as a result.

Conclusions

This mixed methods process evaluation demonstrated a high level of acceptability of all interventions but a strong participant preference for the web-based intervention. Although the web-based intervention was the most preferred, fidelity was lower, and dropout was higher (although more participants actively dropped out of the phone intervention). Despite the high rate of acceptability of the interventions, refinement of the delivery model appeared preferable to some participants. However, any potential modifications to existing interventions should ensure that outcomes are not compromised. The results of this study highlight the strengths and weaknesses of these remotely delivered interventions and offer important aspects for policy makers and practitioners to consider along with the main study outcomes.

Acknowledgments

This research was funded by the New South Wales Ministry of Health Translational Research Grant Scheme. The authors would like to thank the local health district health promotion staff involved in the study for their assistance with recruitment and the parents who participated in the study. They would also like to thank Research Forum Consulting for conducting the qualitative telephone interviews, Christophe Lecathelinais for conducting the randomization, and the Computer-Assisted Telephone Interview team at the University of Newcastle for conducting the telephone calls.

XSL•FO
RenderX

Authors' Contributions

LW, ADO, RAJ, RJW, SE, SY, AG, CR, CIH, VL, and MLH conceived and designed the study; MLH, FS, RJW, JX, and JJ coordinated data collection; MLH conducted the statistical analyses; MLH prepared the original draft of the manuscript; RJW, RAJ, ADO, LW, SY, AG, JX, SE, JJ, FS, CR, and MLH contributed to writing, review, and editing. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Local health district staff recruitment survey.
[DOCX File , 14 KB - jmir_v24i5e35771_app1.docx]

Multimedia Appendix 2
Process evaluation questions.
[DOCX File , 22 KB - jmir_v24i5e35771_app2.docx]

Multimedia Appendix 3
Qualitative telephone interview questions.
[DOCX File , 17 KB - jmir_v24i5e35771_app3.docx]

References

1. National Health Survey: first results, 2017-2018. Australian Bureau of Statistics. 2018. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release#:~:text=In%202017%2D18%2C%20two%20thirds,63.4%25%20in%202014%2D15.&text=There%20was%20a%20large%20increase,46.0%25%20in%202017%2D18 [accessed 2022-04-13]
2. Natale RA, Messiah SE, Asfour L, Uhlhorn SB, Delamater A, Arheart KL. Role modeling as an early childhood obesity prevention strategy: effect of parents and teachers on preschool children's healthy lifestyle habits. J Dev Behav Pediatr 2014;35(6):378-387. [doi: 10.1097/DBP.0000000000000074] [Medline: 25007060]
3. Grimes-Robison C, Evans RR. Benefits and barriers to medically supervised pediatric weight-management programs. J Child Health Care 2008 Dec;12(4):329-343. [doi: 10.1177/1367493508096319] [Medline: 19052190]
4. Fitch A, Fox C, Bauerly K, Gross A, Heim C, Judge-Dietz J, et al. Prevention and management of obesity for children and adolescents. Institute for Clinical Systems Improvement. 2013. URL: https://jesse.tg/ngc-archive/summary/10019#:~:text=Healthy%20Diet&text=Whole%20milk%20is%20recommended%20for,%25)%20milk%20should%20be%20used [accessed 2022-04-13]
5. Wyse R, Wolfenden L, Campbell E, Campbell KJ, Wiggers J, Brennan L, et al. A cluster randomized controlled trial of a telephone-based parent intervention to increase preschoolers' fruit and vegetable consumption. Am J Clin Nutr 2012 Jul;96(1):102-110 [FREE Full text] [doi: 10.3945/ajcn.111.030585] [Medline: 22623749]
6. 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] [Medline: 30735139]
7. 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] [Medline: 30868745]
8. Hammersley ML, Wyse RJ, Jones RA, Wolfenden L, Yoong S, Stacey F, et al. Translation of two healthy eating and active living support programs for parents of 2-6 year old children: a parallel partially randomised preference trial protocol (the 'time for healthy habits' trial). BMC Public Health 2020 May 07;20(1):636 [FREE Full text] [doi: 10.1186/s12889-020-08526-7] [Medline: 32381052]
9. Hammersley ML, Wyse RJ, Jones RA, Stacey F, Okely AD, Wolfenden L, et al. Translation of two healthy eating and active living support programs for parents of 2-6-year-old children: outcomes of the 'time for healthy habits' parallel partially randomised preference trial. Nutrients 2021 Sep 24;13(10):3348 [FREE Full text] [doi: 10.3390/nu13103348] [Medline: 34684348]
10. Schlechter CR, Rosenkranz RR, Guagliano JM, Dzewaltowski DA. A systematic review of children's dietary interventions with parents as change agents: application of the RE-AIM framework. Prev Med 2016 Oct;91:233-243. [doi: 10.1016/j.ypmed.2016.08.030] [Medline: 27569830]

XSL•FO
RenderX

11. Gillespie J, Hughes AR, Gibson AM, Haines J, Taveras EM, Stewart L, et al. Healthy Habits, Happy Homes Scotland (4HS) feasibility study: translation of a home-based early childhood obesity prevention intervention evaluated using RE-AIM framework. Public Health Pract 2020 Nov;1:100026. [doi: 10.1016/j.puhip.2020.100026]
12. Croyden DL, Vidgen HA, Esdaile E, Hernandez E, Magarey A, Moores CJ, et al. A narrative account of implementation lessons learnt from the dissemination of an up-scaled state-wide child obesity management program in Australia: PEACH™ (Parenting, Eating and Activity for Child Health) Queensland. BMC Public Health 2018 Mar 13;18(1):347 [FREE Full text] [doi: 10.1186/s12889-018-5237-8] [Medline: 29534700]
13. Welsby D, Nguyen B, O'Hara BJ, Innes-Hughes C, Bauman A, Hardy LL. Process evaluation of an up-scaled community based child obesity treatment program: NSW Go4Fun®. BMC Public Health 2014 Feb 10;14:140 [FREE Full text] [doi: 10.1186/1471-2458-14-140] [Medline: 24512080]
14. Preference Collaborative Review Group. Patients' preferences within randomised trials: systematic review and patient level meta-analysis. BMJ 2008 Oct 31;337:a1864 [FREE Full text] [doi: 10.1136/bmj.a1864] [Medline: 18977792]
15. Ainsworth HR, Torgerson DJ, Kang 'Ombe AR. Conceptual, design, and statistical complications associated with participant preference. Ann Am Acad Political Soc Sci 2010 Feb 23;628(1):176-188. [doi: 10.1177/0002716209351524]
16. Bradley-Gilbride J, Bradley C. Partially randomized preference trial design. In: Salkind NJ, editor. Encylopedia of Research Design. Thousand Oaks, CA, USA: Sage Publications; 2010.
17. Younge JO, Kouwenhoven-Pasmooij TA, Freak-Poli R, Roos-Hesselink JW, Hunink MM. Randomized study designs for lifestyle interventions: a tutorial. Int J Epidemiol 2015 Dec;44(6):2006-2019. [doi: 10.1093/ije/dyv183] [Medline: 26377509]
18. Brewin CR, Bradley C. Patient preferences and randomised clinical trials. BMJ 1989 Jul 29;299(6694):313-315 [FREE Full text] [doi: 10.1136/bmj.299.6694.313] [Medline: 2504416]
19. Torgerson DJ, Sibbald B. Understanding controlled trials. What is a patient preference trial? BMJ 1998 Jan 31;316(7128):360 [FREE Full text] [doi: 10.1136/bmj.316.7128.360] [Medline: 9487173]
20. 2016 Census Quickstats. Australian Bureau of Statistics. Canberra, Australia: Australian Bureau of Statistics; 2016. URL: https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/036#:~:text=In%20the%202016%20Census%2C%20there,up%202.8%25%20of%20the%20population.&text=The%20median%20age%20of%20people%20in%20Australia%20was%2038%20years [accessed 2022-04-13]
21. Adult literacy and life skills survey. Australian Bureau of Statistics. 2006. URL: https://www.abs.gov.au/ausstats/abs@.nsf/dossbytitle/2D7F8204FEA1D124CA2572E9008079F1?OpenDocument [accessed 2022-04-13]
22. Taki S, Russell CG, Lymer S, Laws R, Campbell K, Appleton J, et al. A mixed methods study to explore the effects of program design elements and participant characteristics on parents' engagement with an mHealth program to promote healthy infant feeding: the Growing Healthy Program. Front Endocrinol (Lausanne) 2019 Jun 25;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]
23. Bentley GF, Jago R, Turner KM. Mothers' perceptions of the UK physical activity and sedentary behaviour guidelines for the early years (Start Active, Stay Active): a qualitative study. BMJ Open 2015 Sep 08;5(9):e008383 [FREE Full text] [doi: 10.1136/bmjopen-2015-008383] [Medline: 26351186]
24. Moore SA, Faulkner G, Rhodes RE, Brussoni M, Chulak-Bozzer T, Ferguson LJ, et al. Impact of the COVID-19 virus outbreak on movement and play behaviours of Canadian children and youth: a national survey. Int J Behav Nutr Phys Act 2020 Jul 06;17(1):85 [FREE Full text] [doi: 10.1186/s12966-020-00987-8] [Medline: 32631350]
25. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008 Jun;41(3-4):327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
26. Ekambareshwar M, Taki S, Mihrshahi S, Baur LA, Rissel C, Wen LM. Participant experiences of an infant obesity prevention program delivered via telephone calls or text messages. Healthcare (Basel) 2020 Mar 16;8(1):60 [FREE Full text] [doi: 10.3390/healthcare8010060] [Medline: 32188139]
27. 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]
28. Jelalian E, Evans EW, Rancourt D, Ranzenhofer L, Taylor N, Hart C, et al. JOIN for ME: testing a scalable weight control intervention for adolescents. Child Obes 2020 Apr;16(3):192-203 [FREE Full text] [doi: 10.1089/chi.2019.0053] [Medline: 31855057]
29. Hardy LL, Mihrshahi S, Gale J, Nguyen B, Baur LA, O'Hara BJ. Translational research: are community-based child obesity treatment programs scalable? BMC Public Health 2015 Jul 14;15:652 [FREE Full text] [doi: 10.1186/s12889-015-2031-8] [Medline: 26169687]
30. Cui Z, Seburg EM, Sherwood NE, Faith MS, Ward DS. Recruitment and retention in obesity prevention and treatment trials targeting minority or low-income children: a review of the clinical trials registration database. Trials 2015 Dec 10;16:564 [FREE Full text] [doi: 10.1186/s13063-015-1089-z] [Medline: 26651822]
31. Laws R, Campbell KJ, van der Pligt P, Russell G, Ball K, Lynch J, et al. The impact of interventions to prevent obesity or improve obesity related behaviours in children (0-5 years) from socioeconomically disadvantaged and/or indigenous families:

XSL•FO
RenderX

a systematic review. BMC Public Health 2014 Aug 01;14:779 [FREE Full text] [doi: 10.1186/1471-2458-14-779] [Medline: 25084804]
32. Davison KK, Charles JN, Khandpur N, Nelson TJ. Fathers' perceived reasons for their underrepresentation in child health research and strategies to increase their involvement. Matern Child Health J 2017 Feb;21(2):267-274 [FREE Full text] [doi: 10.1007/s10995-016-2157-z] [Medline: 27473093]
33. Morgan PJ, Young MD. The influence of fathers on children's physical activity and dietary behaviors: insights, recommendations and future directions. Curr Obes Rep 2017 Sep;6(3):324-333. [doi: 10.1007/s13679-017-0275-6] [Medline: 28762103]
34. Morgan PJ, Young MD, Lloyd AB, Wang ML, Eather N, Miller A, et al. Involvement of fathers in pediatric obesity treatment and prevention trials: a systematic review. Pediatrics 2017 Feb;139(2):e20162635 [FREE Full text] [doi: 10.1542/peds.2016-2635] [Medline: 28130430]

Abbreviations

ECEC: Early Childhood Education and Care
LHD: local health district
NSW: New South Wales
RCT: randomized controlled trial

Edited by R Kukafka; submitted 16.12.21; peer-reviewed by R Kelishadi; comments to author 05.02.22; revised version received 16.02.22; accepted 21.03.22; published 26.05.22.

Please cite as:
Hammersley ML, Wyse RJ, Jones RA, Okely AD, Wolfenden L, Eckermann S, Xu J, Green A, Stacey F, Yoong SL, Jackson J, Innes-Hughes C, Li V, Rissel C
Telephone and Web-Based Delivery of Healthy Eating and Active Living Interventions for Parents of Children Aged 2 to 6 Years: Mixed Methods Process Evaluation of the Time for Healthy Habits Translation Trial
J Med Internet Res 2022;24(5):e35771
URL: https://www.jmir.org/2022/5/e35771
doi:10.2196/35771
PMID:35616994

©Megan L Hammersley, Rebecca J Wyse, Rachel A Jones, Anthony D Okely, Luke Wolfenden, Simon Eckermann, Joe Xu, Amanda Green, Fiona Stacey, Sze Lin Yoong, Jacklyn Jackson, Christine Innes-Hughes, Vincy Li, Chris Rissel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2022. 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.

XSL•FO
RenderX

Original Paper

Disparity and Factors Associated With Internet Health Information Seeking Among US Adults Living With Diabetes Mellitus: Cross-sectional Study

Ransome Eke[1*], MD, MPH, PhD; Khadiza Tul Jannat[2*], MSc; Xin Thomas Yang[3*], MS, PhD; Jason M Parton[3*], PhD

[1]Department of Community Medicine, Mercer University School of Medicine, Columbus, GA, United States

[2]College of Communication and Information Sciences, The University of Alabama, Tuscaloosa, AL, United States

[3]Institute of Data and Analytics, The University of Alabama, Tuscaloosa, AL, United States

[*]all authors contributed equally

Corresponding Author:
Ransome Eke, MD, MPH, PhD
Department of Community Medicine
Mercer University School of Medicine
1633 1st Ave
Columbus, GA, 31901
United States
Phone: 1 7062235235
Email: eke_r@mercer.edu

Abstract

Background: Many patients with chronic medical conditions search the internet to obtain medical advice and health information to improve their health condition and quality of life. Diabetes is a common chronic disease that disproportionately affects different race and ethnicity groups in the United States. In the existing literature on the popularity of internet health information seeking among persons with a chronic medical condition, there are limited data on US adults living with diabetes.

Objective: This study aims to examine the factors associated with internet health information seeking among US adults living with diabetes and whether there is a disparity in internet health information seeking stratified by race and ethnicity.

Methods: We conducted a cross-sectional study using the Health Information National Trends Survey data from 2017 to 2020. We selected our study sample based on respondents' reports on whether they were told they had diabetes, and our primary outcome was internet health information–seeking behavior. We used 2 multivariable logistic regression models to examine the effects of sociodemographic factors and other covariates on the internet health information–seeking behavior of adults with diabetes. Jackknife replicate weights were used to provide bias-corrected variance estimates.

Results: Our study sample included 2903 adults who self-reported that they had diabetes. In total, 60.08% (1744/2903) were non-Hispanic White individuals, 46.88% (1336/2850) were men, and 64% (1812/2831) had some college or graduate education. The prevalence of internet health information seeking in this population was 64.49% (1872/2903), and the main factors associated with internet health information seeking included education level (some college vs less than high school: odds ratio [OR] 1.42, 95% CI 1.44-1.88; and college graduate or higher vs less than high school: OR 2.50, 95% CI 1.79-3.50), age (age group ≥65 years vs age group 18-44 years: OR 0.46, 95% CI 0.34-0.63), and household income level (P<.001). In addition, we found significant differences in the effects of predictors stratified by race.

Conclusions: The findings from this study suggest that internet health information seeking is common among US adults living with diabetes. Internet health information could influence the relationship between health care providers and adults living with diabetes and improve their self-management and quality of life.

(J Med Internet Res 2022;24(5):e32723) doi:10.2196/32723

KEYWORDS

diabetes; internet; information seeking; adults living with diabetes; predictors; racial disparity; Health Information National Trends Survey; race; ethnicity; chronic conditions; self-management; internet health information; mobile phone

XSL•FO
RenderX

Introduction

Background

Health information seeking through internet platforms is increasingly popular [1-3]. An abundance of research has been conducted to explore why people look for health information, what types of health information they seek, how it influences individuals' behaviors, and who are more likely to seek health information on web-based platforms [3-7]. Commonly, people search for health information using internet technology to access relevant health information outside a health care facility. In addition, a positive connection between sociodemographic status and the frequency of health information seeking using the internet has been established [8]. For instance, underprivileged groups are more likely to use web-based health information than the majority groups [9], and being younger and female has been found to be a consistent predictor of eHealth use [10]. Other factors linked to the frequency of web-based health information seeking include income, sex, race and ethnicity, age, and the exposure level of an individual [1,11].

Diabetes mellitus is a common chronic medical condition that disproportionately affects the US adult population. Results obtained from the 2011 to 2016 National Health and Nutrition Examination Surveys data indicated that the prevalence of total diabetes among adult non-Hispanic White individuals with diagnosed diabetes was approximately 12%. In non-Hispanic Black and Hispanic individuals, the prevalence was approximately 20% and 22%, respectively [12]. Studies show that diabetes mellitus is a major chronic disease that carries a significant socioeconomic burden, and the prevalence is projected to rise in the future [13-15]. Management of this condition requires high-quality clinical care and self-management to reduce the risk of associated complications and improve quality of life [11]. Behavior modification and self-management are crucial in effectively managing persons living with diabetes.

Pattern of Health Information Seeking Among Persons With Diabetes

Research suggests that information accessibility is an efficient tool and support necessary to improve chronic medical conditions, including diabetes [16]. Few studies have described the pattern of health information seeking among persons with diabetes. Studies suggest that persons with diabetes have been engaged in passive or active information-seeking activities [17-19]. Passive information-seeking activities involve reading the newspaper and watching television, whereas active information seeking involves mainly using the internet as a source of health information [17,18]. Morgan and Trauth [20] used the Integrated Model of E-Health Use developed by Dutta-Bergman [21] to investigate eHealth information–seeking behavior among persons with diabetes in Greece. The authors [20] found that people with diabetes exhibited different health information–searching behavior because of the intrinsic motivation resulting from access to health care providers or resources.

There is a shortage of data on the internet health information–seeking pattern among adults with diabetes in the United States. Given the growing popularity of internet health information–seeking behavior and the differences in the prevalence of diabetes in the United States, it is vital to understand the factors that predict the use of the internet to seek health information among US adults with diabetes. In addition, with the reported disproportionate racial prevalence of diabetes in the United States, it is essential to investigate whether there is a racial or ethnic disparity in internet health information seeking. Knowing this information is critical for improving diabetes health education and communication, support systems, and quality of life of adults with diabetes in the United States. This study examines the factors associated with internet health information seeking and racial disparity in internet health information seeking among US adults with diabetes.

Methods

Data Source

This cross-sectional study uses data from the Health Information National Trends Survey (HINTS). HINTS is a national representative survey that collects data from the US noninstitutionalized adult population [22]. Conducted by the National Cancer Institute, the survey assesses trends in health information seeking, health information technology adoptions, health communication, knowledge, attitudes, and behavior.

To identify our study population, we pooled and combined data from 4 administrations of HINTS: 2017 (version 5, cycle 1, N=3285), 2018 (version 5, cycle 2, N=3504), 2019 (version 5, cycle 3, N=3374), and 2020 (version 5, cycle 4, N=3865). This study focused on investigating internet health information seeking among the adult population with diabetes. We selected respondents who answered "Yes" to the question "Has a doctor or other health professional ever told you that you had diabetes or high blood sugar?" A total of 2903 respondents met the inclusion criteria for this study (655/3285, 19.94%, in 2017; 714/3504, 20.38%, in 2018; 717/3374, 21.25%, in 2019; and 817/3865, 21.14%, in 2020).

Ethics Approval

This study was approved as exempt by the institutional review board of the University of Alabama because no human participants were involved.

Dependent Variable

Our dependent variable, internet health information–seeking behavior, was defined on the basis of the respondents' report on whether they had in the past 12 months used a computer, smartphone, or other electronic means to look for health or medical information for themselves (yes or no). We excluded invalid or missing responses (52/2903, 1.79%) in our final analyses because the percentage was very small.

Predictor Variables

The primary predictor variables of interest in this study included sociodemographic information: race and ethnicity (non-Hispanic White, non-Hispanic Black, and other), sex (male and female), age group (18-44 years, 45-64 years, and ≥65 years), education

XSL•FO
RenderX

level (less than high school, high school graduate, some college, and college graduate or higher), occupation (employed and unemployed), household income (<US $50,000, US $50,000 to <US $75,000, and ≥US $75,000), residency (urban and rural), and marital status (married, divorced, widowed, single, or never been married). Other covariates included were frequency of visits to health care providers (≤1 time, 2-4 times, and ≥5 times), insurance type (private, public, mixed, no insurance, and other), quality of care (excellent or very good, good, and fair or poor), general health (excellent or very good, good or fair, and poor), ability to take care of one's health (completely or very confident, somewhat confident, and a little or not confident at all). We also examined the respondents' level of trust in the different sources of information (medical professionals, internet, social network, traditional media, and organizations). The trust scores were reverse coded: 4=a lot, 3=some, 2=a little, and 1=or not at all. Medical professionals as a source of information was scored using only 1 question: "From a doctor?" The social network score was based on the mean of 2 questions: "From family or friends?" and "From religious organizations or leaders?" The internet score was based on 1 question: "Internet?" The traditional media score was based on the mean of 2 questions: "From radio?" and "From television?" The newspapers and magazines score was based on 1 question: "From newspapers or magazines?" The trust in organizations score was based on the mean of 2 questions: "From government health agencies?" and "From charitable organizations?"

Statistical Analysis

We used descriptive analyses to summarize the frequencies and unweighted and weighted proportions of respondents grouped by sociodemographic characteristics. The weighted proportions were generated using the survey's weighting variables to generalize the results to the US population. We calculated the trust score using the original survey questions and estimated the mean trust scores for the different sources of information. Multivariable logistic regression models were created to explore the association between the independent variables and health information–seeking behaviors. A total of 2 multiple logistic regression models were constructed to determine the impact of variables of interest with covariates (model 1) and without covariates (model 2). Jackknife replicate weights were used to provide bias-corrected variance estimates [22]. All analyses were conducted using SAS software (version 9.4; SAS Institute Inc), and $P<.05$ was considered statistically significant.

Results

Weighted and Unweighted Estimates

The weighted and unweighted estimates of the characteristics of interest are summarized in Table 1. A total of 2903 respondents with self-reported diabetes were selected. In total, 60.08% (1744/2903) of the respondents were non-Hispanic White individuals, and male respondents accounted for 46.88% (1336/2850) of the total samples. A little more than half of the respondents were aged ≥65 years (1463/2903, 50.4%), and 64% (1812/2831) had some college or graduate education. Most of the respondents lived in urban areas (2519/2903, 86.77%) or reported having an annual household income of <US $50,000 (1460/2568, 56.85%). Figure 1 shows that medical professionals were the most trusted among all sources of health information, meaning the scores were not significantly different across races ($P=.12$). In addition to medical professionals, patients (or people) with diabetes also trust the internet and organizations, and no significant differences were found among the 3 racial groups. The trust in traditional media on health information was lowest in each racial group compared with the trust in other sources. The trust in traditional media was significantly lower in the non-Hispanic White group than in the non-Hispanic Black and other groups ($P<.001$). Trust in social networks, newspapers, and magazines was also not different among the different racial groups (non-Hispanic White, non-Hispanic Black, and other groups). Table 2 presents the characteristics of the respondents who searched the internet for health information compared with those who did not search the internet for health information. Overall, 61.76% (1793/2903) reported that they searched the internet for health information for themselves. Among those who responded yes (1793/2903, 61.76%) to whether they used the internet for health information, most were women (959/1769, 54.21%), non-Hispanic White individuals (1118/1793, 62.35%), residing in an urban area (1595/1793, 88.96%), and married (998/1766, 56.51%). There were significant differences in age group ($P<.001$), education level ($P<.001$), occupation ($P=.01$), sex ($P=.03$), and household income ($P<.001$). In addition, we observed a significant relationship between marital status ($P<.001$) and insurance types of respondents ($P=.002$) and the internet health information seeking among other races with diabetes.

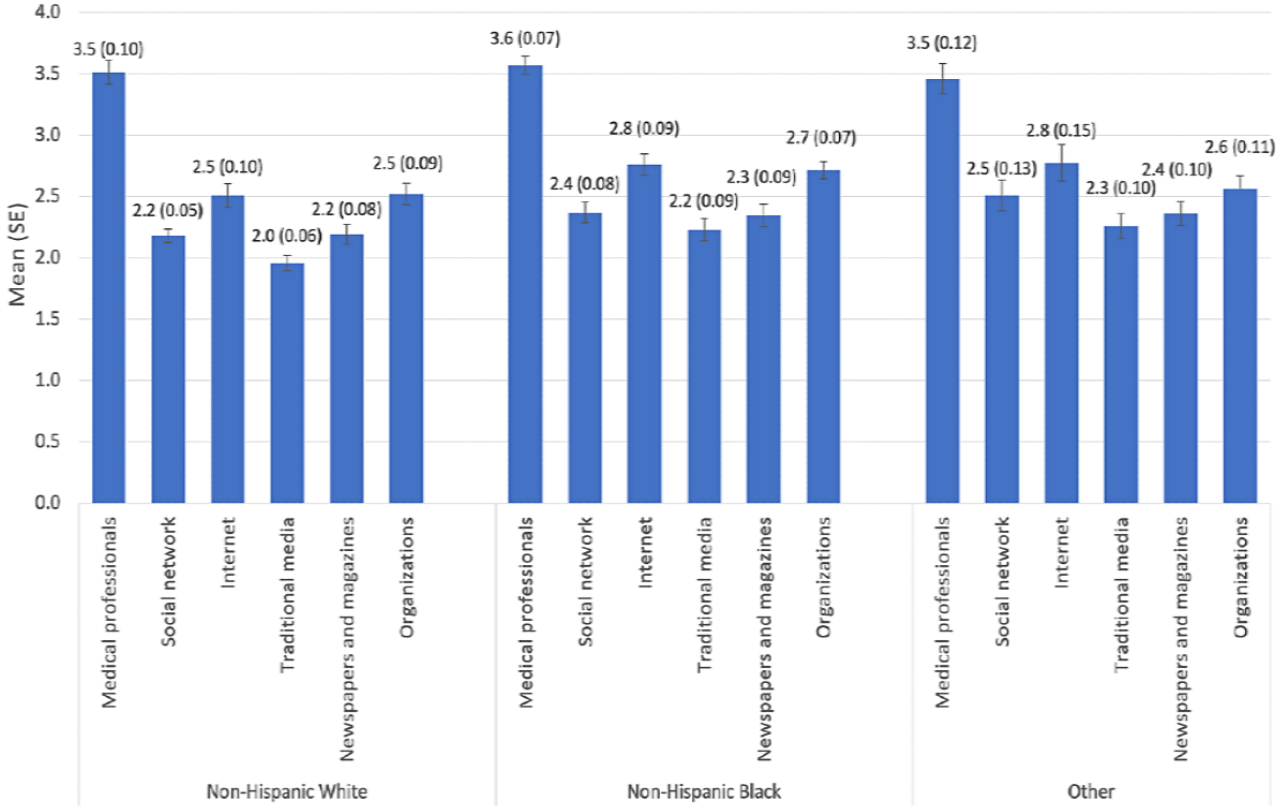
XSL•FO
RenderX

Table 1. Sample characteristics of respondents with diabetes, of Health Information National Trends Survey, 2017 to 2020 (N=2903).

Variable	Value, n	Unweighted estimates (%)	Weighted estimates (%)
Race and ethnicity			
Non-Hispanic White	1744	60.08	65.96
Non-Hispanic Black	638	22	16.8
Other	521	17.9	17.3
Sex			
Male	1336	46.88	48.91
Female	1514	53.12	51.09
Age group (years)			
18 to 44	306	10.5	17.8
45 to 64	1134	39.06	48.96
≥65	1463	50.40	33.25
Education			
Less than high school	337	11.9	13.8
High school graduate	682	24.1	28.5
Some college	908	32.1	37.8
College graduate or higher	904	31.9	19.9
Residency			
Urban	2519	86.77	84.84
Rural	384	13.2	15.2
Marital status			
Married	1426	50.46	57.38
Divorced	575	20.3	12.1
Widowed	433	15.3	8.6
Single or never been married	392	13.9	21.9
Household income (US $)			
<50,000	1460	56.85	54.02
50,000 to <75,000	429	16.7	17.4
≥75,000	679	26.4	28.6

XSL•FO
RenderX

Figure 1. Mean scores of trust in health information sources stratified by race and ethnicity group.



XSL•FO
RenderX

Table 2. Comparison of characteristics between Health Information National Trends Survey (2017 to 2020) respondents who reported that they searched the internet for health information and those who did not search the internet for health information (N=2903).

Variable	Searched the internet for health information, n (%)		*P* value
	Yes (n=1793)	No (n=1110)	
Age group (years)			<.001
18 to 44	226 (12.6)	80 (7.21)	
45 to 64	798 (44.51)	336 (30.27)	
≥65	769 (42.89)	694 (62.52)	
Race and ethnicity			.41
Non-Hispanic White	1118 (62.35)	626 (56.4)	
Non-Hispanic Black	354 (19.74)	284 (25.59)	
Other	321 (17.9)	200 (18.02)	
Education			<.001
Less than high school	106 (6)	231 (21.73)	
High school graduate	316 (17.87)	366 (34.43)	
Some college	616 (34.84)	292 (27.47)	
College graduate or higher	730 (41.29)	174 (16.37)	
Occupation			.01
Employed	355 (19.8)	104 (9.37)	
Unemployed	1438 (80.2)	1006 (90.63)	
Household income (US $)			<.001
<50,000	789 (48.02)	671 (72.54)	
50,000 to <75,000	300 (18.26)	129 (13.95)	
≥75,000	554 (33.72)	125 (13.51)	
Sex			.03
Male	810 (45.79)	526 (48.66)	
Female	959 (54.21)	555 (51.34)	
Residency			.15
Urban	1595 (88.96)	924 (83.24)	
Rural	198 (11.04)	186 (16.76)	
Marital status			<.001
Married	998 (56.51)	428 (40.38)	
Divorced	334 (18.91)	241 (22.73)	
Widowed	180 (10.19)	253 (23.87)	
Single or never been married	254 (14.38)	138 (13.02)	
Insurance type			.001
Private	566 (31.57)	173 (15.59)	
Public	546 (30.45)	521 (46.94)	
Mixed	540 (30.12)	333 (30)	
Other	39 (2.18)	20 (1.8)	
No insurance	102 (5.69)	63 (5.68)	
Quality of care			.60
Excellent or very good	1224 (73.82)	729 (75.86)	
Good	331 (19.96)	164 (17.07)	

XSL•FO
RenderX

Variable	Searched the internet for health information, n (%)		*P* value
	Yes (n=1793)	No (n=1110)	
Fair or poor	103 (6.21)	68 (7.08)	
General health			.29
Excellent or very good	479 (26.88)	272 (24.66)	
Good	1215 (68.18)	746 (67.63)	
Fair or poor	88 (4.94)	85 (7.71)	
Ability to take care of health			.05
Completely or very confident	1071 (59.87)	691 (62.7)	
Somewhat confident	579 (32.36)	304 (27.59)	
A little or not confident at all	139 (7.77)	107 (9.71)	

Factors Associated With the Use of the Internet to Seek Health Information

The 2 models we created to examine the factors that predict the use of the internet to seek health information in participants who reported that they had diabetes are presented in Table 3. In model 1 (the effect of the main predictors without covariates), the results showed that a higher education level was significantly associated with internet health information seeking compared with a less than high school education level (some college: odds ratio [OR] 1.42, 95% CI 1.44-1.88; college graduate or higher: OR 2.50, 95% CI 1.79-3.50). The older age group (≥65 years) was less likely to seek internet health information than the younger age group (18 to 44 years; OR 0.46, 95% CI 0.34-0.63). In addition, income level significantly predicted the use of the internet to seek health information among respondents with self-reported diabetes. Respondents with a household income of ≥US $75,000 had 40% higher odds of searching for health information on the internet than those with a household income of <US $50,000 (OR 1.43, 95% CI 1.03-1.99). Furthermore, we found that respondents who had made ≥5 visits to health care providers during the past 12 months were significantly more likely to use the internet to seek health information than those with fewer visits (OR 1.52, 95% CI 1.11-2.10).

Model 2 presents the effect of the main predictor variables of interest without the covariates. In this model, only education level, age group, and household income level remained the main predictors of use of the internet to search for health information among persons who reported that they had diabetes. In models 1 and 2, race and ethnicity, occupation, sex, marital status, and urbanity were not significantly associated with using the internet to search for health information among respondents who have diabetes (*P*>.05).

Table 4 shows the multivariable analyses of our main predictor variables and covariates on internet health information seeking stratified by race and ethnicity groups (non-Hispanic White, non-Hispanic Black, and other races). We observed significant differences among the several factors across different race and ethnicity groups. In the non-Hispanic White group, the respondents who seek health information on the internet were more likely to have college graduate or higher degrees, ≥5 visits to health care providers, and no insurance, although older age was significantly associated with lower odds of seeking health information on the internet (OR 0.53, 95% CI 0.34-0.82). Among the non-Hispanic Black respondents, individuals who use the internet to seek health information were more likely to have a household income of US $75,000, whereas those living in a rural area (OR 0.40, 95% CI 0.25-0.66) and the older age group (OR 0.32, 95% CI 0.14-0.72) were associated with lower odds of using the internet to seek health information. Among respondents who were neither non-Hispanic White nor non-Hispanic Black in terms of race and ethnicity, those seeking health information on the internet were more likely to have some college degree, have a household income between US $50,000 and US $75,000, live in an urban area, and be widowed.

XSL•FO
RenderX

Table 3. Odds ratios (ORs) and 95% CIs of respondents seeking health information by multiple logistic regression model.

Variable	Model 1, OR (95% CI)	Model 2, OR (95% CI)
Education		
Less than high school (reference)	—[a]	—
High school graduate	0.82 (0.61-1.10)	0.82 (0.60-1.10)
Some college	1.44 (1.10-1.88)[b]	1.45 (1.12-1.86)[b]
College graduate or higher	2.50 (1.79-3.50)[b]	2.76 (2.04-3.74)[b]
Occupation		
Employed (reference)	—	—
Unemployed	0.99 (0.73-1.33)	0.98 (0.79-1.23)
Age group (years)		
18 to 44 (reference)	—	—
45 to 64	0.88 (0.63-1.22)	0.94 (0.72-1.22)
≥65	0.46 (0.34-0.63)[b]	0.46 (0.35-0.60)[b]
Household income (US $)		
<50,000 (reference)	—	—
50,000 to <75,000	0.82 (0.63-1.08)	0.88 (0.69-1.11)
≥75,000	1.43 (1.03-1.99)[b]	1.41 (1.07-1.87)[b]
Residency		
Urban (reference)	—	—
Rural	0.77 (0.59-1.00)	0.81 (0.65-1.02)
Marital status		
Married (reference)	—	—
Divorced	0.85 (0.61-1.19)	0.87 (0.66-1.15)
Widowed	0.85 (0.59-1.22)	0.89 (0.62-1.27)
Single, never been married	0.92 (0.58-1.44)	0.87 (0.58-1.30)
Sex		
Male (reference)	—	—
Female	1.12 (0.94-1.34)	1.15 (0.99-1.34)
Race and ethnicity		
Non-Hispanic White (reference)	—	—
Non-Hispanic Black	1.12 (0.79-1.59)	1.05 (0.80-1.39)
Other	0.95 (0.61-1.46)	0.92 (0.67-1.26)
Insurance type		
Private (reference)	—	N/A[c]
Public	0.88 (0.54-1.45)	N/A
Mixed	1.25 (0.79-2.00)	N/A
Other	0.54 (0.15-1.98)	N/A
No insurance	1.40 (0.55-3.54)	N/A
Frequency of visits to health care providers		
≤1 time (reference)	—	N/A
2 to 4 times	0.88 (0.63-1.22)	N/A
≥5 times	1.52 (1.11-2.10)[b]	N/A

XSL•FO
RenderX

Variable	Model 1, OR (95% CI)	Model 2, OR (95% CI)
Quality of care		
Excellent or very good (reference)	—	N/A
Good	1.16 (0.80-1.69)	N/A
Fair or poor	1.02 (0.68-1.52)	N/A
General health		
Excellent or very good (reference)	—	N/A
Good or fair	1.06 (0.82-1.37)	N/A
Poor	0.70 (0.43-1.14)	N/A
Ability to take care of health		
Completely or very confident (reference)	—	N/A
Somewhat confident	1.21 (0.89-1.64)	N/A
A little or not confident at all	0.97 (0.62-1.52)	N/A

[a]Reference level for corresponding predictors.

[b]*P* values met the threshold for statistical significance.

[C]N/A: not applicable (variables were included in model 1 only).

XSL•FO
RenderX

Table 4. Odds ratios (ORs) and 95% CIs of respondents seeking health information on the internet by race and ethnicity group.

Variable	Model 1: non-Hispanic White, OR (95% CI)	Model 2: non-Hispanic Black, OR (95% CI)	Model 3: other, OR (95% CI)
Education			
Less than high school (reference)	__[a]	—	—
High school graduate	0.70 (0.49-1.01)	1.23 (0.58-2.61)	0.57 (0.20-1.58)
Some college	1.31 (0.95-1.79)	1.36 (0.76-2.45)	4.73 (2.23-10.01)[b]
College graduate or higher	2.77 (1.93-3.96)[b]	1.72 (0.73-4.03)	2.10 (0.53-8.26)
Occupation			
Employed (reference)	—	—	—
Unemployed	0.97 (0.69-1.36)	0.90 (0.50-1.63)	0.70 (0.35-1.41)
Age group (years)			
18 to 44 (reference)	—	—	—
45 to 64	0.81 (0.52-1.26)	1.05 (0.46-2.39)	1.33 (0.49-3.59)
≥65	0.53 (0.34-0.82)[b]	0.32 (0.14-0.72)[b]	0.30 (0.09-1.04)
Household income (US $)			
<50,000 (reference)	—	—	—
50,000 to <75,000	0.93 (0.64-1.36)	0.74 (0.36-1.53)	0.35 (0.18-0.70)[b]
≥75,000	1.47 (0.97-2.24)	2.42 (1.09-5.38)[b]	1.37 (0.74-2.53)
Sex			
Male (reference)	—	—	—
Female	1.06 (0.85-1.33)	1.18 (0.74-1.88)	1.95 (0.90-4.22)
Frequency of visits to health care providers			
≤1 time (reference)	—	—	—
2 to 4 times	1.03 (0.67-1.56)	0.62 (0.33-1.17)	0.85 (0.30-2.37)
≥5 times	1.87 (1.25-2.80)[b]	0.84 (0.42-1.66)	1.48 (0.46-4.74)
Insurance type			
Private (reference)	—	—	—
Public	0.71 (0.37-1.37)	0.69 (0.36-1.33)	2.76 (0.50-15.09)
Mixed	1.00 (0.53-1.89)	1.77 (0.68-4.59)	1.24 (0.29-5.36)
Other	0.53 (0.12-2.43)	1.63 (0.27-9.88)	0.06 (0.01-0.52)[b]
No insurance	2.63 (1.19-5.81)[b]	0.42 (0.07-2.57)	2.67 (0.52-13.69)
Residency			
Urban (reference)	—	—	—
Rural	0.91 (0.68-1.20)	0.40 (0.25-0.66)[b]	0.22 (0.06-0.77)[b]
Marital status			
Married (reference)	—	—	—
Divorced	1.02 (0.70-1.47)	0.69 (0.38-1.24)	1.25 (0.38-4.14)
Widowed	0.93 (0.58-1.49)	1.28 (0.66-2.45)	0.22 (0.07-0.70)[b]
Single or never been married	0.76 (0.45-1.29)	1.24 (0.62-2.48)	0.68 (0.21-2.24)
Quality of care			
Excellent or very good (reference)	—	—	—

XSL•FO
RenderX

Variable	Model 1: non-Hispanic White, OR (95% CI)	Model 2: non-Hispanic Black, OR (95% CI)	Model 3: other, OR (95% CI)
Good	1.11 (0.70-1.74)	1.64 (0.75-3.58)	0.70 (0.23-2.16)
Fair or poor	1.27 (0.69-2.32)	0.58 (0.25-1.35)	0.90 (0.18-4.43)
General health			
Excellent or very good (reference)	—	—	—
Good	0.92 (0.68-1.24)	1.22 (0.63-2.36)	1.51 (0.54-4.27)
Fair or poor	0.71 (0.40-1.26)	1.35 (0.35-5.13)	0.29 (0.05-1.82)
Ability to take care of health			
Completely or very confident (reference)	—	—	—
Somewhat confident	1.27 (0.86-1.87)	1.46 (0.69-3.09)	1.10 (0.41-2.93)
A little or not confident at all	0.82 (0.50-1.36)	0.76 (0.21-2.71)	3.16 (0.52-19.41)

[a]Reference level for corresponding predictors.
[b]$P<.05$.

Discussion

Principal Findings

Diabetes self-management skills refer to the tasks the patient must carry out to manage or reduce the impact of diabetes on their health status and daily living. The internet is a popular platform where individuals with chronic medical conditions obtain information or opinions to improve their health conditions. This cross-sectional study examined the factors associated with internet health information–seeking behavior among US adults with diabetes. We found that approximately two-thirds of the individuals who reported that they are living with diabetes seek personal health information using the internet. Standard features of the US adults with diabetes who seek internet health information include non-Hispanic White race, some college or graduate-level education, unemployment, being married, women, and living in urban areas. The significant predictors of internet use for health information are education level, age, household income, and frequency of visits to health care provider. Our results show that persons with college graduate–level education or higher have 2.5 times higher odds of seeking health information from the internet than individuals with less than high school education. People with diabetes who frequently visit health care providers (≥5 times per year) are 1.5 times more likely to seek health information from the internet than those who see their provider once or not at all in a year. Older age groups (≥65 years) are significantly less likely to use the internet for health information than younger age groups. We observed inconsistencies, by race, in the factors associated with internet health information seeking among US adults with diabetes. The main predictors of internet health information seeking among non-Hispanic White individuals are college graduate education or higher degree, younger age, no insurance, and higher frequency of visits to health care providers. By contrast, among non-Hispanic Black individuals, the main predictors are higher household income, residency, and age of patients.

The ever-growing availability of the internet increases its utility for accessing health information, especially among people with chronic medical conditions. Even so, health care professionals remain the most trusted source of health information and are trailed by internet sources. As in most studies, we observed that the trust in health information sources among US adults with diabetes was higher for health care professionals than for internet sources [19,23,24]. Even so, our study shows that a large proportion of US adults living with diabetes seek health information using internet sources. This finding supports the high rate of reported internet health information–seeking behavior among persons with chronic medical conditions. Data show that >50% of the adults living with chronic medical conditions have accessed the internet for health information related to their situation. Furthermore, 36% reported that information obtained from the internet was helpful regarding medical advice and health information [25].

Contrary to our findings of a large proportion of US adults with diabetes seeking internet health information, Kalanzi et al [24], in their study conducted in Greece, observed a low ranking in the utility of the internet among their study participants. The differential health information–seeking behavior observed between these 2 populations could be explained by their intrinsic motivations, such as access to health care providers and available resources. Nevertheless, our study suggests that people living with diabetes are becoming better informed and better understand their health problems because of their increasing internet health information–seeking behavior. In addition, this study provides essential information to improve the relationship between health care providers and persons living with diabetes. Establishing a good relationship will improve the management and quality of life of individuals living with diabetes. It is vital for health care providers to actively engage with persons living with diabetes in the decision-making process when caring for them. In addition, health care providers should consider discussing available internet-based resources in their management plan to enhance the use of appropriate resources and accuracy of diabetes health information obtained from the internet source.

Individual characteristics (eg, income, sex, race and ethnicity, age, and education) influence internet health

XSL•FO
RenderX

information–seeking behaviors, regardless of the types of illnesses [9,26-28]. There is a positive relationship between individuals with chronic diseases and the frequency of internet health information seeking, which influences their health behavior changes [29]. Our study found that age, education, and household income were significant factors influencing internet health information seeking among adults living with diabetes mellitus. This finding supports the existing literature [30]. Trust in the source of information influences the connection between age and internet health information seeking. For instance, older people, compared with the young generation, find their physicians or health care providers to be reliable sources for seeking health information compared with internet use [31]. In addition, technology adaptation and trust intersect regarding internet health information seeking between young and older adults. Compared with older adults, younger adults are more likely to adopt the internet to search for health information and to trust health information found on the internet because technology adaptation enables them to differentiate between websites that contain low-quality health information and those that contain high-quality health information [31]. These findings emphasize the connection between technology adaptation or acceptance and eHealth literacy regarding health information seeking [31]. Furthermore, levels of education make a huge difference in internet health information seeking, as described in several studies [8,10]. For instance, individuals with higher levels of education are more likely to seek internet health information than those with lower education. We observed 2.5 times the odds of internet health information seeking in adults with diabetes and some college education or higher degree compared with those with less than high school education. This observation underscores the significance of the role of social determinants in promoting health and health equity for all [32] because "social determinants of health are the conditions in which people are born, grow, live, work and age that shape health" [33].

Overall, our study showed a significant association between higher internet use for health information and higher education levels in all race categories. We did not find any significant association between the race of a person living with diabetes and internet health information–seeking behavior. However, we observed inconsistency in the predictors of internet health information seeking across racial groups of adults living with diabetes. Although no association was observed between insurance types and internet use among the Hispanic and non-Hispanic Black individuals with diabetes, our results show that non-Hispanic White individuals with diabetes who have no insurance are significantly more likely to use the internet for health information than non-Hispanic White individuals with private insurance. Previous reports of the association between internet use and insurance status are mixed. Research mostly shows that people with private insurance are more likely to use the internet to seek health information, which could be attributed to their socioeconomic status [34-36]. However, our findings support a previous report that uninsured persons with a reported chronic medical condition were more likely than those with private insurance to search the internet for health information [37]. This finding could imply that higher internet health information–seeking behavior among persons without health insurance and who have a chronic medical condition such as diabetes may be due to barriers in accessing health services because of their insurance status. Research also shows that individuals who have easy access to health information through their health care providers are less likely to search the internet for health information because they have better access to health care services [38].

In comparison, among non-Hispanic Black individuals with diabetes, the main significant predictors of internet health information seeking include higher household income and living in an urban area. Notably, our study explored the difference in the effects of the predictor on internet health information seeking stratified by race and ethnicity among US adults with diabetes. We were unable to compare our data with any similar studies. However, studies have shown a vast racial divide in internet health information–seeking behavior [29,39-41]. These studies indicate that non-Hispanic Black individuals seek more internet health information than non-Hispanic Black and other races to obtain personal health information and medical advice. For instance, Lorence et al [39], in their study, observed a significant gap in the access to the internet between non-Hispanic White and minority races, with the non-Hispanic White group having more access to the internet for health information than the non-Hispanic Black and Hispanic groups [39]. Further research is needed to explore further the coeffect of race and predictors of internet health information–seeking behavior among US adults with diabetes.

The findings from our study add significantly to the literature; however, the study is not without limitations. First, the data used in this study, HINTS data, are self-report secondary survey data. Therefore, there may be issues with validity and bias in the information collected in this survey. For example, the identification of persons with diabetes is based on the information provided by the respondents. We could not verify this information by using clinical data to determine whether diabetes was diagnosed clinically in these respondents. In addition, the response to our dependent variable could have been overreported or underreported. Second, the HINTS data are cross-sectional data. We could not ascertain the trend in internet information seeking in this population and examine any behavior change during the study period. Third, our analytical approach may be subject to robustness issues related to sample sizes. The small sample size of non-Hispanic Black and other race strata compared with the non-Hispanic White group could have affected our findings in this study. Our pooled approach and use of jackknife weights in our analyses helped minimize potential sampling biases and enhance the generalizability of our results. Even with these limitations, the nationwide sampling approach of the survey data is a great strength of this study.

Conclusions

Our study provides insights into the predictors of internet health information–seeking behavior of US adults living with diabetes. Seeking internet health information is common among adults living with diabetes. To improve the self-management and quality of life of individuals living with diabetes, it is crucial for health care providers to educate patients about reliable and verifiable internet health information sources.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
2. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the Internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
3. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
4. Jiang S, Liu PL. Digital divide and Internet health information seeking among cancer survivors: a trend analysis from 2011 to 2017. Psychooncology 2020 Jan;29(1):61-67. [doi: 10.1002/pon.5247] [Medline: 31652360]
5. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785 [FREE Full text] [doi: 10.1080/23311886.2017.1302785]
6. Zhao Y, Zhang J. Consumer health information seeking in social media: a literature review. Health Info Libr J 2017 Dec;34(4):268-283 [FREE Full text] [doi: 10.1111/hir.12192] [Medline: 29045011]
7. 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]
8. Song L, Chang TY. Do resources of network members help in help seeking? Social capital and health information search. Soc Netw 2012 Oct;34(4):658-669 [FREE Full text] [doi: 10.1016/j.socnet.2012.08.002]
9. Mesch G, Mano R, Tsamir J. Minority status and health information search: a test of the social diversification hypothesis. Soc Sci Med 2012 Sep;75(5):854-858. [doi: 10.1016/j.socscimed.2012.03.024] [Medline: 22633160]
10. 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 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
11. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and sources of health information: the impact of the Internet and its implications for health care providers: findings from the first Health Information National Trends Survey. Arch Intern Med 2005 Dec;165(22):2618-2624. [doi: 10.1001/archinte.165.22.2618] [Medline: 16344419]
12. Cheng YJ, Kanaya AM, Araneta MR, Saydah SH, Kahn HS, Gregg EW, et al. Prevalence of diabetes by race and ethnicity in the United States, 2011-2016. JAMA 2019 Dec 24;322(24):2389-2398 [FREE Full text] [doi: 10.1001/jama.2019.19365] [Medline: 31860047]
13. National Diabetes Statistics Report, 2020: Estimates of Diabetes and Its Burden in the United States. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [accessed 2021-08-07]
14. Lin J, Thompson TJ, Cheng YJ, Zhuo X, Zhang P, Gregg E, et al. Projection of the future diabetes burden in the United States through 2060. Popul Health Metr 2018 Jun 15;16(1):9 [FREE Full text] [doi: 10.1186/s12963-018-0166-4] [Medline: 29903012]
15. Agha M, Agha R. The rising prevalence of obesity: part A: impact on public health. Int J Surg Oncol (N Y) 2017 Aug;2(7):e17 [FREE Full text] [doi: 10.1097/IJ9.0000000000000017] [Medline: 29177227]
16. Shaw RJ, Johnson CM. Health information seeking and social media use on the Internet among people with diabetes. Online J Public Health Inform 2011;3(1):ojphi.v3i1.3561 [FREE Full text] [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
17. Meyfroidt S, Aeyels D, Van Audenhove C, Verlinde C, Peers J, Panella M, et al. How do patients with uncontrolled diabetes in the Brussels-Capital Region seek and use information sources for their diet? Prim Health Care Res Dev 2013 Jul;14(3):229-239. [doi: 10.1017/S1463423612000205] [Medline: 22717540]
18. Longo DR, Schubert SL, Wright BA, LeMaster J, Williams CD, Clore JN. Health information seeking, receipt, and use in diabetes self-management. Ann Fam Med 2010;8(4):334-340 [FREE Full text] [doi: 10.1370/afm.1115] [Medline: 20644188]
19. Kuske S, Schiereck T, Grobosch S, Paduch A, Droste S, Halbach S, et al. Diabetes-related information-seeking behaviour: a systematic review. Syst Rev 2017 Oct 24;6(1):212 [FREE Full text] [doi: 10.1186/s13643-017-0602-8] [Medline: 29065919]
20. Janice Morgan A, Trauth EM. Socio-economic influences on health information searching in the USA: the case of diabetes. Inf Technol People 2013;26(4):324-346 [FREE Full text] [doi: 10.1108/itp-09-2012-0098]
21. Dutta-Bergman MJ. Media use theory and internet use for health care. In: Murero M, Rice RE, editors. The Internet and Health Care: Theory, Research, and Practice. Milton Park, UK: Routledge; 2006:83-103.
22. Hesse BW, Moser RP, Rutten LJ, Kreps GL. The health information national trends survey: research from the baseline. J Health Commun 2006;11 Suppl 1:vii-xvi. [doi: 10.1080/10810730600692553] [Medline: 16641070]

XSL•FO
RenderX

23. Sayakhot P, Carolan-Olah M. Sources of information on Gestational Diabetes Mellitus, satisfaction with diagnostic process and information provision. BMC Pregnancy Childbirth 2016 Sep 29;16(1):287 [FREE Full text] [doi: 10.1186/s12884-016-1067-9] [Medline: 27679990]
24. Kalantzi S, Kostagiolas P, Kechagias G, Niakas D, Makrilakis K. Information seeking behavior of patients with diabetes mellitus: a cross-sectional study in an outpatient clinic of a university-affiliated hospital in Athens, Greece. BMC Res Notes 2015 Feb 20;8:48 [FREE Full text] [doi: 10.1186/s13104-015-1005-3] [Medline: 25889724]
25. Fox S, Purcell K. Chronic disease and the Internet. Pew Research Center. 2010 Mar 24. URL: https://www.pewresearch.org/internet/2010/03/24/chronic-disease-and-the-internet/ [accessed 2022-05-17]
26. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest 2014 Nov;146(5):1387-1394. [doi: 10.1378/chest.14-0970] [Medline: 25367475]
27. Schroeder S, Capocelli KE, Masterson JC, Harris R, Protheroe C, Lee JJ, et al. Effect of proton pump inhibitor on esophageal eosinophilia. J Pediatr Gastroenterol Nutr 2013 Feb;56(2):166-172 [FREE Full text] [doi: 10.1097/MPG.0b013e3182716b7a] [Medline: 23325438]
28. Seamans MJ, Robinson WR, Thorpe Jr RJ, Cole SR, LaVeist TA. Exploring racial differences in the obesity gender gap. Ann Epidemiol 2015 Jun;25(6):420-425 [FREE Full text] [doi: 10.1016/j.annepidem.2015.03.010] [Medline: 25887701]
29. Baker L, Wagner TH, Singer S, Bundorf MK. Use of the Internet and e-mail for health care information: results from a national survey. JAMA 2003 May 14;289(18):2400-2406. [doi: 10.1001/jama.289.18.2400] [Medline: 12746364]
30. Shimabukuro TT, Grosse SD, Rice C. Medical expenditures for children with an autism spectrum disorder in a privately insured population. J Autism Dev Disord 2008 Mar;38(3):546-552. [doi: 10.1007/s10803-007-0424-y] [Medline: 17690969]
31. Miller LM, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
32. Plough A, Miller C, Tait M. Moving forward together: an update on building and measuring a culture of health. Princet Robert Wood Johnson Found. 2018. URL: https://www.rwjf.org/en/library/research/2018/05/moving-foward-together--an-update-on-building-and-measuring-a-culture-of-health.html [accessed 2021-07-28]
33. Artiga S, Hinton E. Beyond health care: the role of social determinants in promoting health and health equity. Kaiser Family Foundation. 2018 May 10. URL: https://www.kff.org/racial-equity-and-health-policy/issue-brief/beyond-health-care-the-role-of-social-determinants-in-promoting-health-and-health-equity/ [accessed 2022-05-17]
34. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California health interview survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
35. Saulsberry L, Price M, Hsu H. Medicare Medicaid Res Rev 2014;4(4):mmrr2014-004-04-b01 [FREE Full text] [doi: 10.5600/mmrr.004.04.b01] [Medline: 25383242]
36. Zimmerman MS, Shaw Jr G. Health information seeking behaviour: a concept analysis. Health Info Libr J 2020 Sep;37(3):173-191. [doi: 10.1111/hir.12287] [Medline: 32052549]
37. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the internet for health information? Health Serv Res 2006 Jun;41(3 Pt 1):819-836 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00510.x] [Medline: 16704514]
38. Swoboda CM, Van Hulle JM, McAlearney AS, Huerta TR. Odds of talking to healthcare providers as the initial source of healthcare information: updated cross-sectional results from the Health Information National Trends Survey (HINTS). BMC Fam Pract 2018 Aug 29;19(1):146 [FREE Full text] [doi: 10.1186/s12875-018-0805-7] [Medline: 30157770]
39. Lorence DP, Park H, Fox S. Racial disparities in health information access: resilience of the Digital Divide. J Med Syst 2006 Aug;30(4):241-249. [doi: 10.1007/s10916-005-9003-y] [Medline: 16978003]
40. Miller EA, West DM, Wasserman M. Health information Websites: characteristics of US users by race and ethnicity. J Telemed Telecare 2007;13(6):298-302. [doi: 10.1258/135763307781644852] [Medline: 17785031]
41. Brodie M, Flournoy RE, Altman DE, Blendon RJ, Benson JM, Rosenbaum MD. Health information, the Internet, and the digital divide. Health Aff (Millwood) 2000;19(6):255-265. [doi: 10.1377/hlthaff.19.6.255] [Medline: 11192412]

Abbreviations

HINTS: Health Information National Trends Survey
OR: odds ratio

XSL•FO
RenderX

Edited by R Kukafka; submitted 08.08.21; peer-reviewed by MDG Pimentel, C Memering; comments to author 30.12.21; revised version received 05.02.22; accepted 06.04.22; published 30.05.22.

Please cite as:
Eke R, Jannat KT, Yang XT, Parton JM
Disparity and Factors Associated With Internet Health Information Seeking Among US Adults Living With Diabetes Mellitus: Cross-sectional Study
J Med Internet Res 2022;24(5):e32723
URL: https://www.jmir.org/2022/5/e32723
doi:10.2196/32723
PMID:35635741

©Ransome Eke, Khadiza Tul Jannat, Xin Thomas Yang, Jason M Parton. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.05.2022. 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.

XSL•FO
RenderX

Original Paper

Comparison Between the 24-hour Holter Test and 72-hour Single-Lead Electrocardiogram Monitoring With an Adhesive Patch-Type Device for Atrial Fibrillation Detection: Prospective Cohort Study

Soonil Kwon[1], MD; So-Ryoung Lee[1], MD, PhD; Eue-Keun Choi[1,2], MD, PhD; Hyo-Jeong Ahn[1], MD; Hee-Seok Song[3], MSc; Young-Shin Lee[3], MSc; Seil Oh[1,2], MD, PhD; Gregory Y H Lip[1,2,4,5], MD

[1]Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[2]Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

[3]Seers Technology Co, Seongnam-si, Republic of Korea

[4]Liverpool Centre for Cardiovascular Science, University of Liverpool, Liverpool, United Kingdom

[5]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark

Corresponding Author:
Eue-Keun Choi, MD, PhD
Department of Internal Medicine
Seoul National University Hospital
101 Daehak-ro
Jongno-gu
Seoul
Republic of Korea
Phone: 82 2 2072 0688
Fax: 82 2 762 9662
Email: choiek417@gmail.com

Abstract

Background: There is insufficient evidence for the use of single-lead electrocardiogram (ECG) monitoring with an adhesive patch-type device (APD) over an extended period compared to that of the 24-hour Holter test for atrial fibrillation (AF) detection.

Objective: In this paper, we aimed to compare AF detection by the 24-hour Holter test and 72-hour single-lead ECG monitoring using an APD among patients with AF.

Methods: This was a prospective, single-center cohort study. A total of 210 patients with AF with clinical indications for the Holter test at cardiology outpatient clinics were enrolled in the study. The study participants were equipped with both the Holter device and APD for the first 24 hours. Subsequently, only the APD continued ECG monitoring for an additional 48 hours. AF detection during the first 24 hours was compared between the two devices. The diagnostic benefits of extended monitoring using the APD were evaluated.

Results: A total of 200 patients (mean age 60 years; n=141, 70.5% male; and n=59, 29.5% female) completed 72-hour ECG monitoring with the APD. During the first 24 hours, both monitoring methods detected AF in the same 40/200 (20%) patients (including 20 patients each with paroxysmal and persistent AF). Compared to the 24-hour Holter test, the APD increased the AF detection rate by 1.5-fold (58/200; 29%) and 1.6-fold (64/200; 32%) with 48- and 72-hour monitoring, respectively. With the APD, the number of newly discovered patients with paroxysmal AF was 20/44 (45.5%), 18/44 (40.9%), and 6/44 (13.6%) at 24-, 48-, and 72-hour monitoring, respectively. Compared with 24-hour Holter monitoring, 72-hour monitoring with the APD increased the detection rate of paroxysmal AF by 2.2-fold (44/20).

Conclusions: Compared to the 24-hour Holter test, AF detection could be improved with 72-hour single-lead ECG monitoring with the APD.

(J Med Internet Res 2022;24(5):e37970) doi:10.2196/37970

XSL•FO
RenderX

KEYWORDS

atrial fibrillation; diagnosis; electrocardiogram; wearable device; health monitoring; Holter; cardiac; arrhythmia; electrocardiogram; ECG; EKG; digital tool; cardiology; patient monitoring; outpatient clinic; cardiac health; diagnostic; patient; clinician; digital health

Introduction

Electrocardiogram (ECG) monitoring is essential for the detection of atrial fibrillation (AF). Although a standard 12-lead ECG can be used to detect AF, its diagnostic effectiveness decreases as the AF burden becomes low and multiple snapshots of 12-lead ECGs or ambulatory ECG monitoring are often required [1,2]. Although numerous handheld or wearable ECG devices are now readily available [3], the Holter test remains the gold standard for ambulatory ECG monitoring. Briefly, the Holter test is usually performed over 24 hours and can record multiple ECG leads. However, in the case of paroxysmal AF, the known AF burden is generally less than 5% [4]. In such cases, more extended ECG monitoring is usually necessary to detect AF.

Recently, adhesive patch-type devices (APDs) have been used to detect AF. Compared to the Holter test, APDs are generally more compact and convenient for patients [5]. APDs also have the advantage of an extended monitoring period for up to several days, depending on the product. Therefore, APDs could be a valuable alternative to the Holter test. However, most APDs monitor single-lead ECG such that they can record ECG signals along a single vector. As a result, there are concerns of over- or under-detection of AF compared to the standard Holter test [6]. Additionally, the diagnostic performance of single-lead ECG monitoring could be suboptimal due to noisy tracings, frequent ectopic beats, or the coexistence of other tachyarrhythmias [7]. Although multiple studies have validated the diagnostic performance of single-lead ECG monitoring with APDs for various cardiac arrhythmias [6,8-11], evidence of direct comparisons between the Holter test and single-lead ECG monitoring with an APD for AF detection remains limited [12].

This study aimed to compare the 24-hour Holter test to 72-hour single-lead ECG monitoring with an APD among patients with AF in routine medical care.

Methods

Ethics Approval

The study protocol was approved by the Seoul National University Hospital Institutional Review Board and adhered to the Declaration of Helsinki revised in 2013 (IRB No: H-2006-224-1138).

Study Design and Population

This was a single-center, prospective cohort study. Among the patients who received outpatient management for AF at our institution (Seoul National University Hospital, Seoul, Republic of Korea), the patients who needed ambulatory ECG monitoring for AF management or evaluation were screened for the study. All patients were medically examined and screened by any of the 3 electrophysiologists (EKC, SRL, or SO). Screening and recruitment processes were conducted in the outpatient clinic setting.

The *inclusion criteria* of the study population were (1) those who were previously diagnosed with AF and (2) those who were indicated for the 24-hour Holter test for routine management or monitoring of AF at outpatient clinics. The *exclusion criteria* were (1) persistent atrial flutter or atrial tachycardia and (2) failure to complete simultaneous single-lead ECG monitoring with the APD and Holter test for the first 24 hours.

Between October 2020 and September 2021, a total of 210 patients were enrolled in the study. Among them, 2 (1%) patients had no AF but persistent atrial tachycardia, and 8 (3.8%) patients did not complete simultaneous monitoring for the first 24 hours due to detachment of the monitoring device or recording errors. Therefore, a total of 200 participants were included in this study.

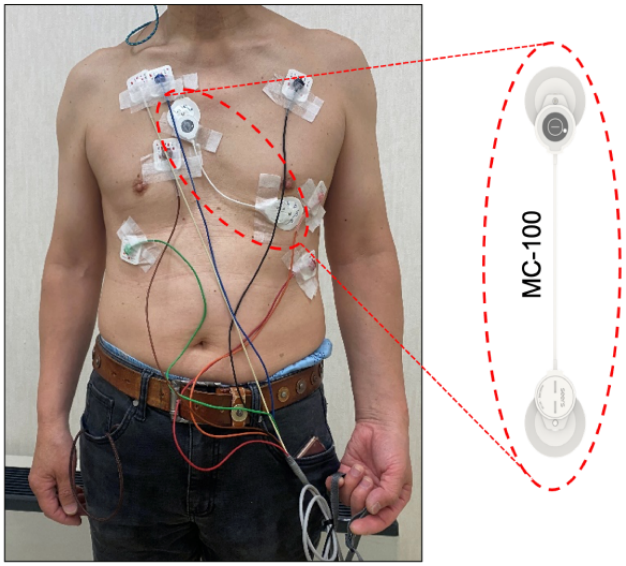
Study Flow

After obtaining informed consent, baseline characteristics were examined by a researcher in the outpatient clinic. Baseline characteristics included demographic information (age, sex, height, body weight, and body mass index), information on AF (types of AF, CHA_2DS_2-VASc [congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category] scores, and history of treatment for AF), comorbidities (hypertension, diabetes mellitus, heart failure, vascular disease, chronic kidney disease, liver disease, and thromboembolism), and concomitant medications (antiarrhythmic agents, diuretics, oral anticoagulants, and antiplatelet agents).

After enrollment, each participant started simultaneous single-lead ECG monitoring and Holter tests for the first 24 hours. An APD (mobiCARE MC-100, Seers technology, Seongnam-si, Gyeonggi-do, Republic of Korea) and a Holter device (SEER Light, GE Healthcare, Chicago, IL, USA) were attached to each participant, as shown in Figure 1. The Holter device was set to record three channels (leads I, V1, and V6), and the electrodes were placed at the positions for standard ECG measurement. The APD was set to record a single channel (lead II) and was placed 45 from the internipple line. Overlap of the electrodes in both devices was avoided to prevent signal noise and interference.

XSL•FO
RenderX

Figure 1. Device setting for electrocardiogram monitoring of a study participant. A study participant recorded a single-lead electrocardiogram (lead II) using an adhesive patch-type device (MC-100) and a three-channel electrocardiogram (lead I, V1, and V6) using the Holter test. Overlap of the electrodes of both devices was avoided to prevent signal noise and interference.



After completing the simultaneous monitoring for 24 hours, participants returned the Holter device and continued single-lead ECG monitoring with the APD for additional 48 hours. After completing the comprehensive monitoring, each participant returned the APD and responded to a survey on the convenience of using the APD. The raw data of both devices were extracted and anonymized to protect participants' privacy. Raw data were independently reviewed and analyzed by 4 cardiologists (SK, SRL, EKC, and HJA). If there was any discrepancy in the interpretation of the ECG signal, the senior electrophysiologist (EKC) decided the final interpretation.

A Brief Specification of MC-100

The APD used in the study (MC-100) has two medical standard 4.0 mm electrode snaps connected by a single wire. The device is compatible with conventional sticky ECG electrodes. It is powered by a commercial CR2032H coin cell battery and can operate continuously for at least 72 hours. The size of the device is 29 mm × 120 mm, and it weighs 8.9 grams. The device can record a single-lead ECG signal with a sampling rate of 256 Hz. Additionally, the device has accelerometers and gyroscopes to measure movement activity. The device is connected to the user's smartphone using Bluetooth and transmits ECG data to the smartphone. A user can access ECG data from a smartphone, and real time monitoring is possible using a preinstalled app. The MC-100 has a built-in memory of 256 kilobytes, which can store ECG data for up to 2-3 minutes if it is disconnected from the smartphone. During validation of the ECG measurements using the MC-100 for the population with non-AF cardiac arrhythmias, the device showed a diagnostic performance comparable to that of a conventional Holter test [13].

Sample Size Determination

We used the McNemar test to estimate the sample size. Based on previous reports [5,8], we assumed that 14% of patients would be negative for the Holter test but positive for the APD due to the extended monitoring period, while 4% of patients would be positive for the Holter test but negative for the APD due to potential disadvantages of single-lead ECG monitoring. To achieve a power of 80% and a two-sided significance of 5%, the study required 194 participants. Considering potential dropouts, we estimated a total of 200 participants are required to conduct the study. The PASS 15 Power Analysis and Sample Size Software was used to perform sample size calculations.

Statistical Analysis

The diagnostic performances of the APD and Holter tests were compared. The variables for the comparison included total monitoring time (minutes), the proportion of noise (ie, uninterpretable portions of the recorded signals, %), AF detection rate (%), and AF burden (%). To compare the variables between the two tests, a paired *t* test or Wilcoxon signed-rank test was performed according to their normality. The AF detection rate and AF burden were measured every 24 hours to observe the diagnostic benefits of extended single-lead ECG monitoring daily. We also recalculated the AF detection rate and AF burden for the APD by only including AF episodes that lasted ≥30 seconds. In all statistical analyses, a *P* value of less than .05 was considered statistically significant. Statistical analyses were performed using SPSS Statistics for Windows, version 22.0 (IBM Corp).

Results

Baseline Characteristics

The baseline characteristics of the study population are presented in Table 1. The mean patient age was 60 years, and 70.5% (141/200) of the patients were male. The proportions of paroxysmal AF and persistent AF were 68% (136/200) and 32% (64/200), respectively. The most common comorbidity was hypertension (54.5%, 109/200). Most participants used beta-blockers (33.5%, 67/200), oral anticoagulants (57%, 114/200), and class Ic antiarrhythmic agents (43%, 86/200).

XSL•FO
RenderX

Table 1. Baseline characteristics of the study population (N=200).

Characteristics	Value
Age (year), mean (SD)	60 (7.8)
Gender, n (%)	
Male	141 (70.5)
Female	59 (29.5)
Intersex, n (%)	0 (0)
Height, cm (SD)	167.2 (7.9)
Weight, kg (SD)	70.2 (11)
Body mass index, kg/m^2 (SD)	25.1 (3.1)
Mean CHA_2DS_2-VASc[a] score (SD)	1.5 (1.1)
Median CHA_2DS_2-VASc score (IQR)	1 (1-2)
AF[b] types, n (%)	
Paroxysmal	136 (68)
Persistent	64 (32)
AF treatment information, n (%)	
Prior electrical cardioversion	53 (26.5)
Prior catheter ablation	122 (61)
Comorbidities, n (%)	
Hypertension	109 (54.5)
Diabetes mellitus	34 (17)
Heart failure	18 (9)
Peripheral artery disease	1 (0.5)
Chronic kidney disease	2 (1)
Chronic liver disease	3 (1.5)
Thromboembolism	2 (1)
Concomitant medications, n (%)	
Beta-blocker	67 (33.5)
Calcium channel blocker	39 (19.5)
RAAS[c] blockade	50 (25)
Diuretics	11 (5.5)
Oral anticoagulant	114 (57)
Antiplatelet agent	25 (12.5)
Class Ic antiarrhythmic agent	86 (43)
Amiodarone	29 (14.5)

[a]CHA_2DS_2-VASc: congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, stroke or transient ischemic attack, vascular disease, age 65 to 74 years, sex category.

[b]AF: atrial fibrillation.

[c]RAAS: renin-angiotensin-aldosterone system.

Comparisons of ECG Monitoring Between the Holter and APD

A total of 200 participants performed ECG monitoring with the Holter and APD. The mean monitoring durations were 1402 (SD 106) min (0.97, SD 0.07 days) and 4242 (SD 401) min (2.95, SD 0.28 days) for the Holter and APD, respectively (*P*<.001; Table 2). The median noise proportions were significantly higher in the APD (median <0.1%, 95% CI 0-0.2 for the Holter; and median 0.3%, 95% CI 0.1-0.7 for the APD,*P*<.001). Most signal noises were caused by motion artifacts, touching of the device, or poor electrode contact. The

XSL•FO
RenderX

APD had additional signal loss due to Bluetooth disconnection from the user's smartphone (median 2%, 95% CI 1.0-4.4).

Table 2. Comparisons of ECG[a] monitoring durations and noise proportions between the Holter and the adhesive patch-type device.

	24-hour Holter monitoring	72-hour single-lead ECG monitoring with an adhesive patch-type device	*P* value
Total participants, N	200	200	N/A[b]
Mean monitoring duration, min (SD)	1402 (106)	4242 (401)	<.001
Median noise proportions, % (IRQ)	<0.1 (0-0.2)	0.3 (0.1-0.7)	<.001

[a]ECG: electrocardiogram.

[b]N/A: not applicable.

Feasibility of 72-Hour ECG Monitoring With the APD

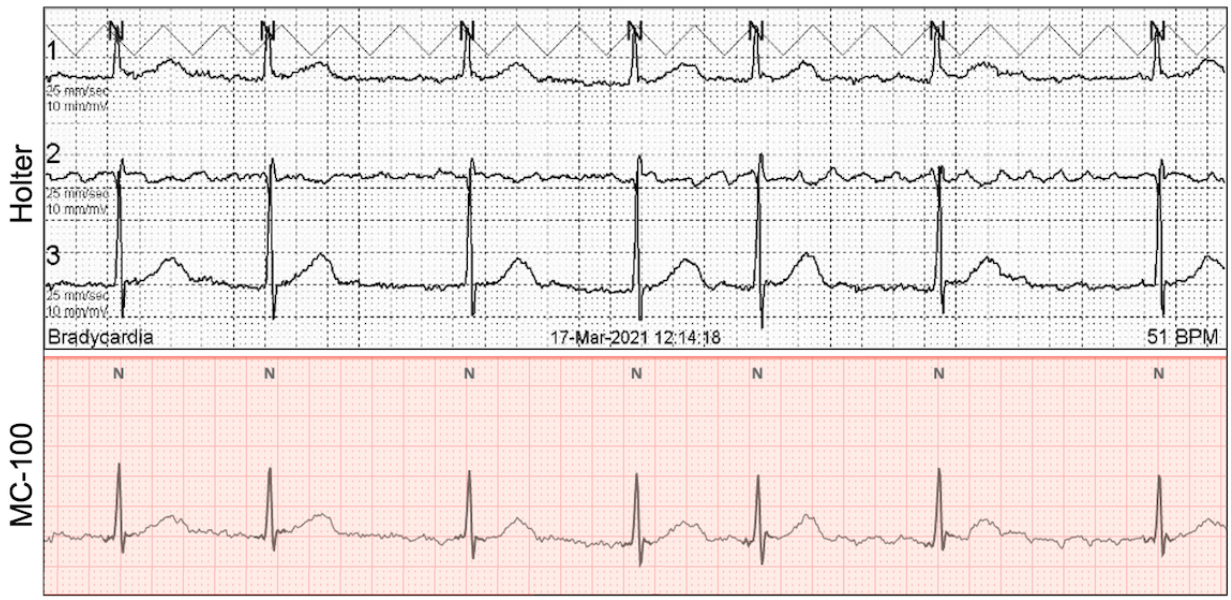
Of the 200 participants, 188 (94%) completed the 72-hour ECG monitoring with the APD. During the extended monitoring period, 12 (6%) participants failed to complete the 72-hour ECG monitoring. Reasons for failing to complete the 72-hour monitoring included device or app errors in 4 (2%) participants, misuse of the device by the user in 3 (1.5%) participants, skin irritation in 2 (1%) participants, and other reasons in 3 (1.5%) participants. The skin irritation that occurred in the 2 participants recovered spontaneously after removing the APD and did not require further medical aid.

Comparisons of AF Detection and AF Burdens Between the Holter and APD

Examples of single-lead ECG monitoring with the APD for persistent and paroxysmal AF are presented in Figure 2 and Figure 3, respectively. For the first 24 hours, both the Holter and APD yielded the same AF detection rate (40/200, 20% of participants; Figure 4). Paroxysmal and persistent AF were equally identified in 20 participants using both devices. During the extended monitoring period, the APD detected paroxysmal AF in 18 (9%) and 6 (3%) new participants on days 2 and 3, respectively. Compared to the 24-hour Holter test, 72-hour ECG monitoring with the APD increased the AF detection rate by 1.6-fold (40/200, 20% with the Holter; and 64/200, 32% with the APD). When comparing only participants with paroxysmal AF, the APD increased the AF detection rate by 2.2-fold (20/180, 11.1% with the Holter; and 44/180, 24.4% with the APD).

The daily distributions of AF burden measured by the 24-hour Holter test and 72-hour single-lead ECG monitoring with the APD are compared in Figure 5. There was no significant difference in the AF burden measured by the two devices on day 1 (*P*=.06). Except for the participants with persistent AF (ie, AF burden of 100%), most AF burdens were less than 5% (165/180, 91.7% on day 1; 163/180, 90.6% on day 2; 156/180, 86.7% on day 3 with the APD). The next most common AF burden was 5%-25% (10/180, 5.5% on day 1; 11/180, 6.1% on days 2 and 3 with the APD). Individual AF burdens changed dynamically over the monitoring period, except in cases of persistent AF (Figure 6).

Figure 2. An example of persistent AF (participant #105) detected by the Holter and adhesive patch-type device. Both the Holter and adhesive patch-type device detected AF coherently. AF: atrial fibrillation.



XSL•FO
RenderX

Figure 3. An example of onset and termination of paroxysmal AF detected by the adhesive patch-type device. Both onset and termination of paroxysmal AF can be accurately detected by the adhesive patch-type device. AF: atrial fibrillation.

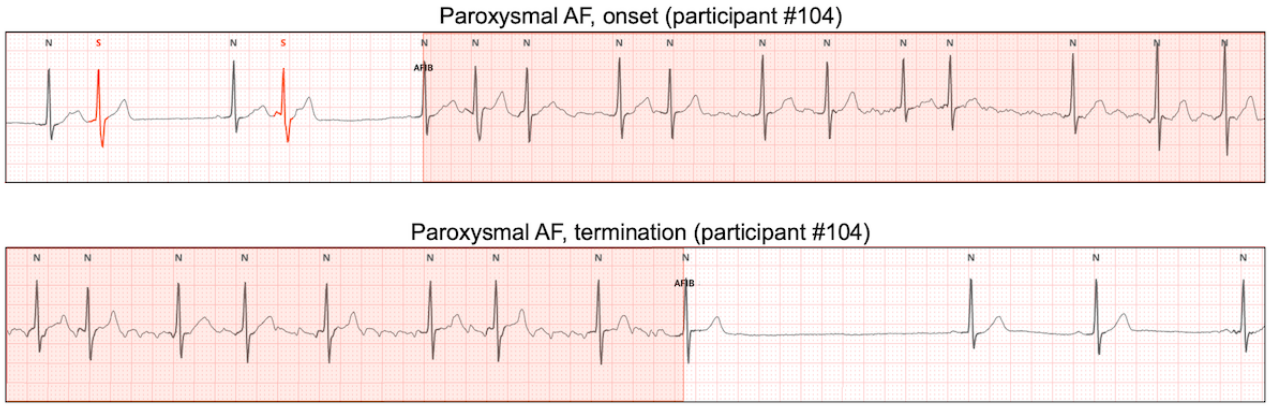


Figure 4. Comparison of AF detection between the Holter and adhesive patch-type device. The daily proportions of participants with AF were detected by the 24-hour Holter test and 72-hour single-lead electrocardiogram monitoring with the adhesive patch-type device. AF: atrial fibrillation.

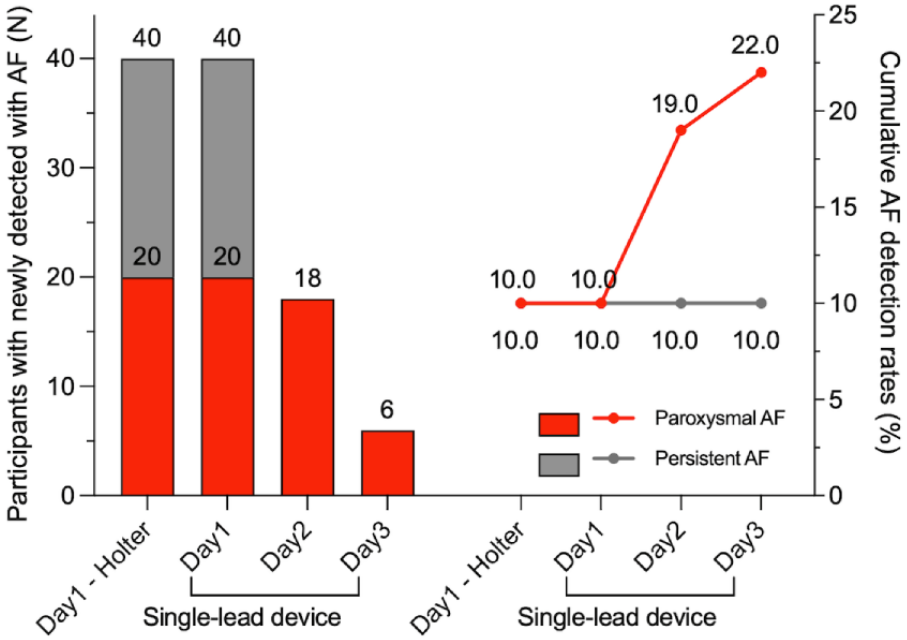
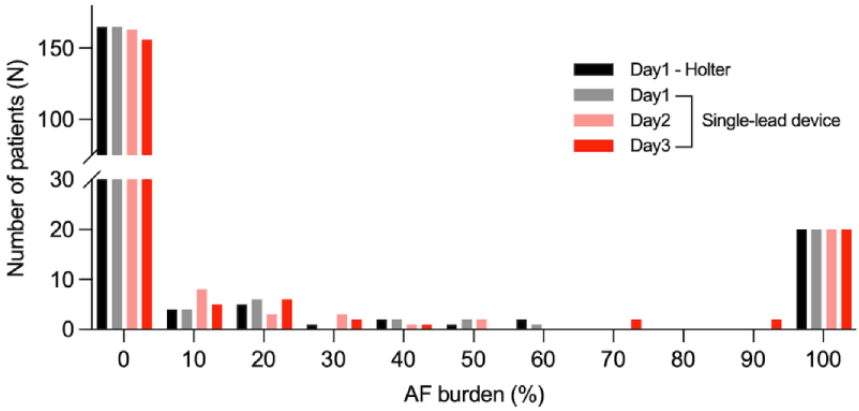
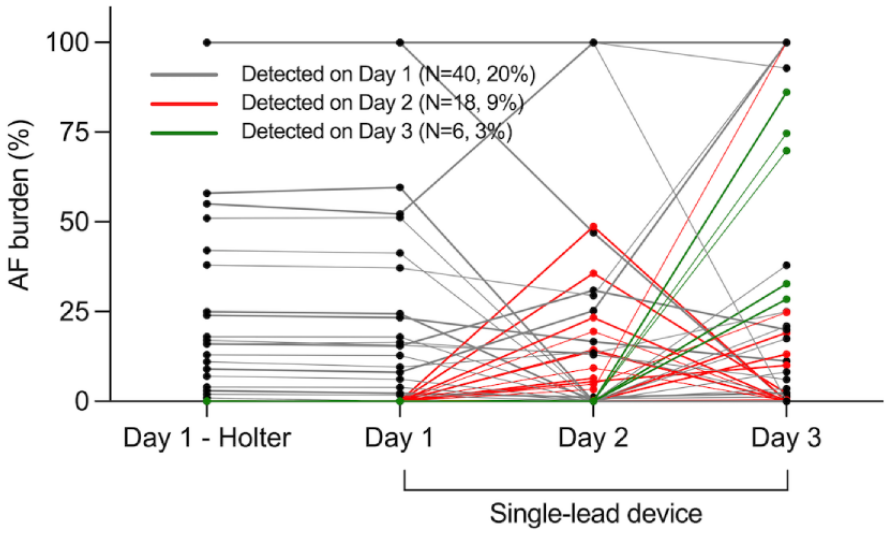


Figure 5. Distribution of AF burden measured by the Holter and adhesive patch-type device. The daily AF burdens were compared between 24-hour Holter monitoring and 72-hour single-lead electrocardiogram monitoring with the adhesive patch-type device. AF: atrial fibrillation.



XSL•FO
RenderX

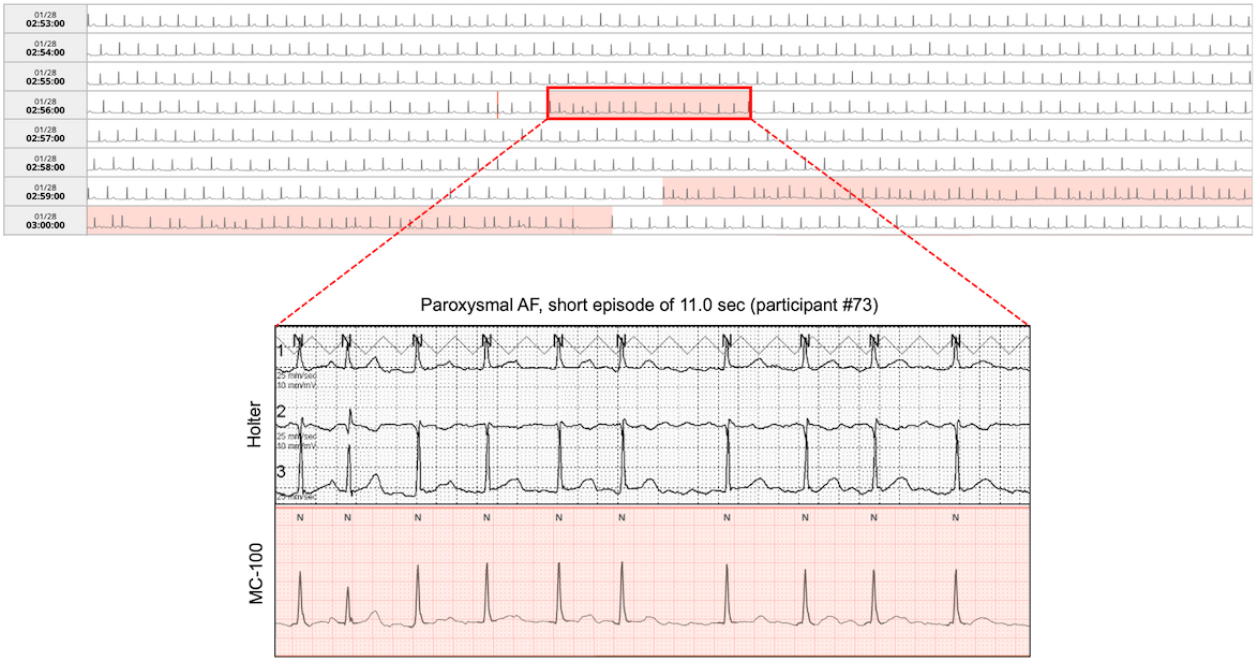
Figure 6. Dynamic changes of daily AF burden. For each participant, daily AF burden was tracked over the monitoring period using the adhesive patch-type device. AF: atrial fibrillation.



Impact of the Duration of AF Episodes on AF Detection

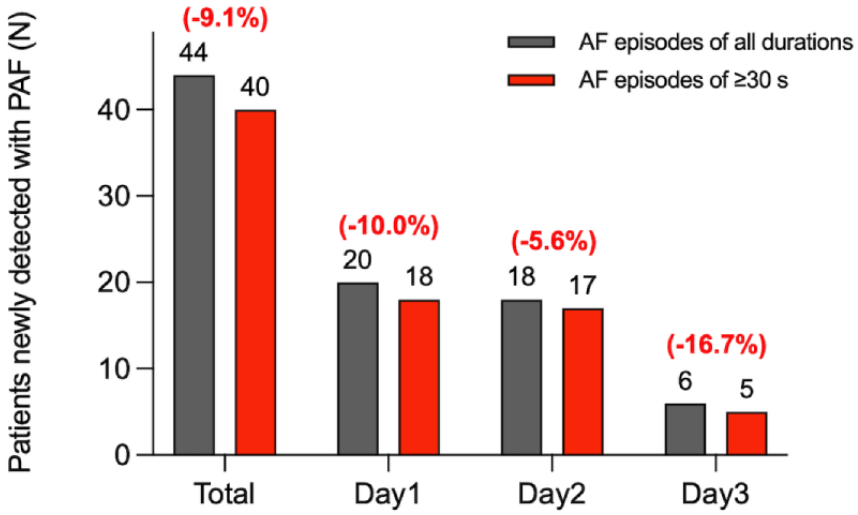
An example of a short episode of paroxysmal AF (duration <30 seconds) is shown in Figure 7. If only episodes lasting over 30 seconds with the APD were counted as AF, the detection rate of paroxysmal AF is presented in Figure 8. Limiting the minimally required duration of AF episodes to 30 seconds decreased the detection rate of paroxysmal AF by 9.1% overall. Despite the decrease in the detection rate of paroxysmal AF, 72-hour single-lead ECG monitoring with the APD yielded a 2-fold higher detection rate than the 24-hour Holter test (20/180, 11.1% with the Holter; and 40/180, 22.2% with the APD).

Figure 7. An example of short episode of paroxysmal AF detected by the Holter and adhesive patch-type device. Both the Holter and adhesive patch-type device detected a short episode of paroxysmal AF accurately. AF: atrial fibrillation.



XSL•FO
RenderX

Figure 8. Comparison of the detection rates of paroxysmal AF with the adhesive patch-type device (MC-100) by episode durations. Limiting the minimally required duration of AF episodes to 30 seconds decreased the detection rate of paroxysmal AF by 9.1% overall. AF: atrial fibrillation; PAF: paroxysmal atrial fibrillation.



The Convenience of Using the APD for Extended ECG Monitoring

The survey results for the use of the APD are presented in Multimedia Appendix 1. The surveys were collected by 190 study participants (190/200, 95%). Most participants did not report discomfort or skin irritability during use of the APD (66/190, 34.7% and 65/190, 34.2%, respectively). Among respondents, 27/190 (14.2%) reported skin irritability. Instances of discomfort using the APD were the most frequent during activity (54/190, 28.4%), followed by that during sleep (35/190, 18.4%). Episodes of intermittent device detachment were observed in 111/190 (58.4%) of the respondents. Overall, more than a half of the respondents responded very positively to using the device and app (107/190, 56.3% and 98/190, 51.6%, respectively).

Discussion

Principal Findings

This study compared AF detection rates with a 24-hour Holter test and 72-hour single-lead ECG monitoring with an APD among patients requiring AF monitoring. The principal findings of the study are as follows: (1) during the simultaneous use of both monitoring methods, they yielded the same AF detection rate over 24 hours; (2) extended monitoring with the APD increased AF detection rates by 1.6-fold compared to those with the 24-hour Holter test (2.2-fold for the case of paroxysmal AF); (3) limiting the minimally required duration of AF episodes to 30 seconds decreased the detection rate of paroxysmal AF by 9.1% with the APD; and (4) most participants responded that it was convenient to use the APD over the extended monitoring period.

The major differences between our study and others are that (1) we compared the AF detection rate between the Holter test and the APD from routine medical care for patients with AF, and (2) we evaluated the impact of the duration of AF episodes on AF detection using the APD.

When managing patients with AF in an outpatient setting, performing a 24-hour Holter test is common. However, we found that the 24-hour Holter test detected only approximately 40/64 (62.5%) of participants with AF compared to those by the 72-hour single-lead ECG monitoring. This result suggests that the outpatient-based 24-hour Holter test is often ineffective for AF detection. Therefore, the management or evaluation of patients with AF might be suboptimal when the 24-hour Holter test is used, especially in cases of paroxysmal AF.

According to recent European guidelines [1], AF-like episodes of at least 30 seconds are required to diagnose AF using a single-lead ECG device. It is challenging to identify P waves as accurately as a standard 12-lead ECG using a single-lead ECG. This study found that identifying AF episodes lasting ≥30 seconds decreased the AF detection rate by 9.1%. Nevertheless, 72-hour single-lead ECG monitoring was superior to that of the 24-hour Holter test for AF detection.

ECG monitoring is an essential method for AF detection. Recently, wearable or portable ECG monitoring devices have become widely accepted for AF detection [14]. Smartwatches with the capability of ECG measurement and APDs are typical examples of newly introduced ECG monitoring tools [3]. Unlike smartwatches or handheld devices, APDs can continuously monitor ECG signals; therefore, APDs have a potential to maximize AF detection rates during a given monitoring period. Some APDs can record multiple-lead ECG, but most devices have been designed to record single-lead ECG to minimize their size and maximize their convenience. The convenience of APDs is that they can monitor ECG signals for an extended period (several days to weeks), and thus increase the possibility of AF detection without disturbing the patient. In addition, they are small and convenient to use [5]; however, one disadvantage is that most APDs can only record single-lead ECG signals. Consequently, if a patient has P waves that are low in amplitude along with the vector between the device's electrodes, there is a possibility of misdiagnosing atrial arrhythmias, including AF [6,13]. However, in this study, both the Holter test and the APD showed equivalent AF detection rates during the first 24 hours.

XSL•FO
RenderX

One possible explanation is that the number of patients might have been insufficient to show the difference in the AF detection rates between the two monitoring methods during the first 24 hours. However, the APD might have been as effective as Holter because the APD detected P waves effectively in our study; the APD's electrodes were attached 120 mm apart and along the P wave axis to increase the detection of P waves. A further study is warranted regarding the effectiveness of P wave detection with the APD.

Some studies investigated the use of APDs for AF detection [6,15,16]. A systematic review by Ramkumar et al [17] found that a moderate linear relationship exists between monitoring time and an AF detection rate for a single-lead ECG device. Although a more extended monitoring with APDs would increase the AF detection rate, the mSToPS trial found that most AF detection occurred within a week [15]. However, as the monitoring period becomes longer, test compliance would decrease while the possibility of skin problems caused by APDs would increase. For example, Heckbert et al [9] reported that APDs with a median monitoring time of 14 days induced skin irritation in 4% of the participants. Similar to the mSToPS trial, this study also found that most AF detection occurred within 7 days. In our study, only 1% (N=2) of the study participants discontinued ECG monitoring with the APD (MC-100) because of skin problems. Skin irritation occurred less commonly in this study than in other studies due to differences in the monitoring period with the APDs [9,10]. In addition, the MC-100 uses conventional ECG snap electrodes that are widely used for ECG measurements, and the contact area between the device and the skin is smaller than that of other commercial products. The smaller contact area of adhesives might also have contributed to a lower prevalence of skin problems in our study. Moreover, the feasibility of extended monitoring with an APD could be an issue due to device detachment during daily activity. However, the APD used in our study was easy to reattach to the body because the device was small and had a simple and lightweight structure. Therefore, in most cases, the detachment period was relatively short. As a result, the proportion of signal noise due to any detachment episodes accounted for only a median of 0.3% (95% CI 0.1-0.7) of the total monitoring time.

Limitations

This study has several limitations. First, it focused on the population diagnosed with AF who received routine medical care at outpatient clinics. Therefore, the diagnostic performance in the general population cannot be estimated. Second, during the extended monitoring period, there were possible false-positive or false-negative episodes with the APD because no Holter data or standard 12-lead ECG was available to validate the episodes. False-positive AF episodes could also be attributed to underdetected ectopic P waves or premature atrial beats with the single-lead ECG data [6,13]. Third, this study cannot determine the AF detection performance of the APD for special cases including concomitant complete atrioventricular block or slow ventricular response as the study participants did not have such cases.

Conclusions

Compared to the 24-hour Holter test, 72-hour single-lead ECG monitoring with an APD could improve AF detection rates. Both tests were equally effective during the first 24 hours despite the potential disadvantages of single-lead ECG monitoring. Focusing on paroxysmal AF, the detection rates could be improved by 2.2-fold with the APD. In addition, the APD was convenient for extended monitoring without causing serious skin irritation. Our results showed that the extended monitoring with the APD for AF detection was feasible and had good compliance. Extended monitoring of single-lead ECG with the APD could be beneficial for AF detection among patients whereby conventional ECG tests were inadequate in documenting AF episodes.

Acknowledgments

This work was supported by the Korea Medical Device Development Fund by the Korean government (Ministry of Science and Information and Communications Technology, Ministry of Trade, Industry and Energy, Ministry of Health and Welfare, Republic of Korea, Ministry of Food and Drug Safety; project number 202013B14) and grant funded by Seers Technology.

Authors' Contributions

SK conceptualized study design and methodology, conducted data curation and database management, performed investigation, formal analysis, and visualization of the study results, and wrote the manuscript. SRL reviewed the methodology, investigation, and data curation, and the manuscript. EKC was the principal investigator of the study; conceptualized and reviewed the study design; reviewed the manuscript; and conducted supervision, project administration, and funding acquisition. HJA conducted data curation and database management and reviewed the manuscript. HSS conducted data curation, performed formal analysis, prepared resources, and reviewed the manuscript. YSL conducted data curation, prepared resources, and reviewed the manuscript. SO reviewed the study design, methodology, and the manuscript. GYHL reviewed the study design, methodology, and the manuscript. All authors have read and agreed to the final manuscript.

Conflicts of Interest

HSH and YSL: stockholders of Seers Technology Co, Ltd. EKC: research grants or speaking fees from Abbott, Bayer, BMS/Pfizer, Biosense Webster, Chong Kun Dang, Daewoong Pharmaceutical Co, Daiichi-Sankyo, DeepQure, Dreamtech Co, Ltd, Jeil Pharmaceutical Co Ltd, Medtronic, Samjinpharm, Seers Technology, and Skylabs. Stock options from Seers Technology, and Skylabs; GYHL: Consultant for Bayer/Janssen, BMS/Pfizer, Medtronic, Boehringer Ingelheim, Novartis, Verseon and

XSL•FO
RenderX

Daiichi-Sankyo. Speaker for Bayer, BMS/Pfizer, Medtronic, Boehringer Ingelheim, and Daiichi-Sankyo. No fees are received personally. SK, SRL, HJA, and SO have none to declare.

Multimedia Appendix 1
The survey on the use of MC-100 (English-translated version).
[DOCX File , 22 KB - jmir_v24i5e37970_app1.docx]

References

1. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2021 Feb 01;42(5):373-498. [doi: 10.1093/eurheartj/ehaa612] [Medline: 32860505]
2. Shen WK, Sheldon RS, Benditt DG, Cohen MI, Forman DE, Goldberger ZD, et al. 2017 ACC/AHA/HRS Guideline for the Evaluation and Management of Patients With Syncope: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. Circulation 2017 Aug 01;136(5):e60-e122. [doi: 10.1161/CIR.0000000000000499] [Medline: 28280231]
3. Sana F, Isselbacher EM, Singh JP, Heist EK, Pathik B, Armoundas AA. Wearable Devices for Ambulatory Cardiac Monitoring: JACC State-of-the-Art Review. J Am Coll Cardiol 2020 Apr 07;75(13):1582-1592 [FREE Full text] [doi: 10.1016/j.jacc.2020.01.046] [Medline: 32241375]
4. Go AS, Reynolds K, Yang J, Gupta N, Lenane J, Sung SH, et al. Association of Burden of Atrial Fibrillation With Risk of Ischemic Stroke in Adults With Paroxysmal Atrial Fibrillation: The KP-RHYTHM Study. JAMA Cardiol 2018 Jul 01;3(7):601-608 [FREE Full text] [doi: 10.1001/jamacardio.2018.1176] [Medline: 29799942]
5. Barrett PM, Komatireddy R, Haaser S, Topol S, Sheard J, Encinas J, et al. Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring. Am J Med 2014 Jan;127(1):95.e11-95.e17 [FREE Full text] [doi: 10.1016/j.amjmed.2013.10.003] [Medline: 24384108]
6. Nault I, André P, Plourde B, Leclerc F, Sarrazin JF, Philippon F, et al. Validation of a novel single lead ambulatory ECG monitor - Cardiostat™ - Compared to a standard ECG Holter monitoring. J Electrocardiol 2019;53:57-63. [doi: 10.1016/j.jelectrocard.2018.12.011] [Medline: 30641305]
7. Pitman BM, Chew SH, Wong CX, Jaghoori A, Iwai S, Thomas G, et al. Performance of a Mobile Single-Lead Electrocardiogram Technology for Atrial Fibrillation Screening in a Semirural African Population: Insights From "The Heart of Ethiopia: Focus on Atrial Fibrillation" (TEFF-AF) Study. JMIR mHealth uHealth 2021 May 19;9(5):e24470 [FREE Full text] [doi: 10.2196/24470] [Medline: 34009129]
8. Liu C, Chang SL, Yeh YH, Chung FP, Hu YF, Chou CC, et al. Enhanced detection of cardiac arrhythmias utilizing 14-day continuous ECG patch monitoring. Int J Cardiol 2021 Jun 01;332:78-84. [doi: 10.1016/j.ijcard.2021.03.015] [Medline: 33727122]
9. Heckbert SR, Austin TR, Jensen PN, Floyd JS, Psaty BM, Soliman EZ, et al. Yield and consistency of arrhythmia detection with patch electrocardiographic monitoring: The Multi-Ethnic Study of Atherosclerosis. J Electrocardiol 2018;51(6):997-1002 [FREE Full text] [doi: 10.1016/j.jelectrocard.2018.07.027] [Medline: 30497763]
10. Karaoğuz MR, Yurtseven E, Aslan G, Deliormanlı BG, Adıgüzel Ö, Gönen M, et al. The quality of ECG data acquisition, and diagnostic performance of a novel adhesive patch for ambulatory cardiac rhythm monitoring in arrhythmia detection. J Electrocardiol 2019;54:28-35. [doi: 10.1016/j.jelectrocard.2019.02.012] [Medline: 30851474]
11. Pradhan S, Robinson JA, Shivapour JK, Snyder CS. Ambulatory Arrhythmia Detection with ZIO® XT Patch in Pediatric Patients: A Comparison of Devices. Pediatr Cardiol 2019 Jun;40(5):921-924. [doi: 10.1007/s00246-019-02089-0] [Medline: 30937502]
12. Giebel GD, Gissel C. Accuracy of mHealth Devices for Atrial Fibrillation Screening: Systematic Review. JMIR mHealth uHealth 2019 Jun 16;7(6):e13641 [FREE Full text] [doi: 10.2196/13641] [Medline: 31199337]
13. Kwon S, Lee SR, Choi EK, Ahn HJ, Song HS, Lee YS, et al. Validation of Adhesive Single-Lead ECG Device Compared with Holter Monitoring among Non-Atrial Fibrillation Patients. Sensors (Basel) 2021 Apr 30;21(9):3122 [FREE Full text] [doi: 10.3390/s21093122] [Medline: 33946269]
14. Bansal A, Joshi R. Portable out-of-hospital electrocardiography: A review of current technologies. J Arrhythm 2018 Apr;34(2):129-138 [FREE Full text] [doi: 10.1002/joa3.12035] [Medline: 29657588]
15. Steinhubl SR, Waalen J, Edwards AM, Ariniello LM, Mehta RR, Ebner GS, et al. Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation: The mSToPS Randomized Clinical Trial. JAMA 2018 Jul 10;320(2):146-155 [FREE Full text] [doi: 10.1001/jama.2018.8102] [Medline: 29998336]
16. Turakhia MP, Ullal AJ, Hoang DD, Than CT, Miller JD, Friday KJ, et al. Feasibility of extended ambulatory electrocardiogram monitoring to identify silent atrial fibrillation in high-risk patients: the Screening Study for Undiagnosed

XSL•FO
RenderX

Atrial Fibrillation (STUDY-AF). Clin Cardiol 2015 May;38(5):285-292 [FREE Full text] [doi: 10.1002/clc.22387] [Medline: 25873476]
17. Ramkumar S, Nerlekar N, D'Souza D, Pol DJ, Kalman JM, Marwick TH. Atrial fibrillation detection using single lead portable electrocardiographic monitoring: a systematic review and meta-analysis. BMJ Open 2018 Sep 17;8(9):e024178 [FREE Full text] [doi: 10.1136/bmjopen-2018-024178] [Medline: 30224404]

Abbreviations

AF: atrial fibrillation
APD: adhesive patch-type device
ECG: electrocardiogram

Edited by T Leung; submitted 14.03.22; peer-reviewed by E van der Velde, H Karim, E Chen; comments to author 05.04.22; revised version received 12.04.22; accepted 23.04.22; published 09.05.22.

Please cite as:
Kwon S, Lee SR, Choi EK, Ahn HJ, Song HS, Lee YS, Oh S, Lip GYH
Comparison Between the 24-hour Holter Test and 72-hour Single-Lead Electrocardiogram Monitoring With an Adhesive Patch-Type Device for Atrial Fibrillation Detection: Prospective Cohort Study
J Med Internet Res 2022;24(5):e37970
URL: https://www.jmir.org/2022/5/e37970
doi:10.2196/37970
PMID:35532989

©Soonil Kwon, So-Ryoung Lee, Eue-Keun Choi, Hyo-Jeong Ahn, Hee-Seok Song, Young-Shin Lee, Seil Oh, Gregory Y H Lip. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Influence of Perceived Social Presence on the Willingness to Communicate in Mobile Medical Consultations: Experimental Study

Lijuan Chen[1*], DPhil; Danyang Zhang[2*], MSc; Mutian Hou[3], MA

[1]Department of Journalism, School of Humanities, Shanghai University of Finance and Economics, Shanghai, China

[2]Department of Journalism and Communication, School of Media and Communication, Shanghai Jiao Tong University, Shanghai, China

[3]Psychological Research and Counseling Center, Southwest Jiaotong University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Danyang Zhang, MSc
Department of Journalism and Communication
School of Media and Communication
Shanghai Jiao Tong University
800 Dongchuan Rd
Shanghai, 200240
China
Phone: 86 138 6752 9369
Email: zdytiffany@163.com

Abstract

Background: With the rise of online health care service, there is growing discussion on the relationship between physicians and patients online, yet few researchers have paid attention to patients' perception of social presence, especially its influence on their willingness to communicate (WTC).

Objective: The goal of the research is to investigate the influence of perceived social presence (PSP) on WTC in mobile medical consultations.

Methods: Participants living in Yunnan province during the period of middle to high risk of COVID-19 infection were recruited via the internet. They were assigned randomly into 2 groups interacting with a virtual physician presenting high and low levels of social presence and then asked to complete a questionnaire. Based on the theoretical framework, the study puts forward a model evaluating the relationships among participants' PSP, communication apprehension (CA), self-perceived communication competence (SPCC), and willingness to communicate about health (WTCH) in the computer-mediated communication between virtual physicians and patients.

Results: In total 206 (106 in group 1 and 100 in group 2) valid samples were gathered (from 276 log-ins) and 88.8% (183/206) of them were aged 18 to 44 years, which approximately resembles the age distribution of the main population engaging in online medical consultation in China. Independent *t* test shows that there is significant difference between the PSP of the 2 groups (P=.04), indicating a successful manipulation of social presence. The total effect of PSP on WTCH is 0.56 (P<.001), among which 74.4% is direct effect (P<.001). Among the indirect effects between PSP and WTCH, the mediating effect of SPCC accounts for 68.8% (P<.001) and the sequential mediating effect of CA→SPCC accounts for 19.2% (P<.001), while the mediating effect of CA alone is not significant (P=.08).

Conclusions: This study provides a comprehensible model, demonstrating that PSP is an important antecedent of WTCH, and the sequential mediating effect of CA and SPCC found in this study also proves that in the environment of online mobile medical services, CA cannot affect communication directly. The findings will provide some practical inspiration for the popularization of online medical service, especially for the promotion of online physician-patient communication.

(J Med Internet Res 2022;24(5):e31797) doi:10.2196/31797

XSL•FO
RenderX

KEYWORDS

mobile medical consultation; mobile medical service; perceived social presence; willingness to communicate about health; COVID-19

Introduction

Background

Physician-patient communication is a hot button issue in China. Web- and mobile phone–based medical consultations as a supplement to traditional outpatient services have gradually become the prevalent pathway [1,2] to consult professionals about health problems, especially during the challenge of COVID-19 [3]. Compared with the traditional way, online health care uses relatively fewer medical resources [1] and is supplementary to the "previous form of health communication...based on face-to-face interpersonal communication and linear traditional media communication" [2]. Therefore, it is worth studying the factors and mechanisms enabling patients to feel comfortable and more willing to communicate with physicians.

Willingness to communicate (WTC) is the possibility that a person chooses to communicate (especially to start a conversation) under the condition that they are free to make a choice [4]. WTC not only depends on the individual's innate personality but the context of the communication as well [5]. Communication plays an important role in the tension between the physician and patient and the patient's family members [6]. "Willingness to communicate about health (WTCH) represents a situational application of the WTC construct" [7]. The way the patient experiences medical care is dynamic and complex. Many factors in the process will moderate the relationship between communication and desired outcomes [8]. Physician-patient communication via the internet may have risks of uncertainty of information quality and medical advices [9]. The patient information-seeking behavior is a 1-way communication, while visiting physicians at a hospital is a kind of 2-way interaction. Patients are usually regarded as passive recipients of the health information, and they do not always actively participate in medical consultations [10,11]. However, the patient's participation in making a medical decision is the key to patient-centered communication [8].

Many researchers have studied WTC between physicians and patients in specific situations such as organ donation [12], clinical trials [13], and sexual health [14], etc. Petrič et al [15] found that social interactions such as communication in online health communities will affect the physician-patient relationship. The willingness to accept online treatment has also been proven to partially depend on WTC online with general practitioners [16]. Published studies have proved that WTC is closely related to the context of interaction [17,18]; however, most of them only verified the correlation, and few have tried to verify the causality. In addition, for a long time communication apprehension (CA) and perceived communication competence were considered the 2 best predictors of WTC [19], but there have not been enough empirical studies to discuss the impact of this mechanism on patients' willingness to talk about health in a computer-mediated environment.

Compared with face-to-face interactions, online physician-patient communication has particularities such as connectivity, textuality, asynchronism, and anonymity [1], which may exert a complicated effect on the online physician-patient communication process due to the lack of social cues. The theory of social presence has been widely applied in research studies of computer-mediated communication (CMC), which lacks social cues. Therefore, to fill the research gap, this study aims to explore the underlying mechanism of the influence of social presence on users' WTC during online mobile medical services by controlled experiments. We hope to make contributions to the understanding of online patient WTC within the framework of social presence theory and shed light on practical implications for online medical platform designers to improve the consultation service.

Theoretical Basis

Short et al [20] define social presence as the extent to which the communication medium facilitates social emotional exchange, and the extent to which a person can experience and understand another person and the underlying personal relationship. This definition is often used to depict the perception of others being there [21,22] and is often associated with concepts such as immediacy [23], intimacy [24], authenticity [25], and social copresence [26], etc. As an inherent property of the communication medium, social presence is related to the medium's capability to convey nonverbal cues [20]. From the perspective of psychology, social presence can be described as the "warmth" of the media—namely, the capability to make people feel human warmth and sensitivity [27].

With the development of CMC, online interactions have become more immersive. Because of the relative absence of social context information and feedback, CMC hinders the transmission of cues such as personality or hints conveyed by nonverbal behaviors in face-to-face interactions [28], which reduces social presence [29]. The perception of social presence is proved to positively influence users' trust and intention in the online environment. For instance, it can increase consumer trust in an online shopping environment, willingness to purchase online [30,31], electronic loyalty [32] and intention of continuous use [33]. Social presence is crucial for inspiring the patient's willingness to use the service, especially in an online health community based on support exchange [34]. Peng et al [35] found that in online patient-to-physician communities, social presence and especially its characteristics that enable users to feel comfortable, safe, and warm with a sense of belonging and sensitivity will positively influence users' information-seeking behaviors and their willingness to participate, which promotes the physician-patient relationship. In general, exploration on the effect of social presence in online health care is still inadequate compared with other fields of online medical research.

Until now, researchers had not reached an agreement on the measurement of perceived social presence. Prior studies mostly

XSL•FO
RenderX

used the self-report questionnaire to measure social presence. Such measurement may show relatively higher levels of significance; however, there may be more interference factors during the process. Rourke et al [36] conducted a content analysis of online discussion. Despite the 12 resulting indicators to measure social presence, they still acknowledged the limitations of developing and testing the efficacy of an instrument to measure social presence. Social presence has been proven to be closely related to the information richness theory [37]. According to the theory, the increment of social cues will increase the capacity of potential information carrying. Hassanein and Head [38] tested and verified the forecasting effect of social presence on user attitudes toward the internet by providing texts and picture design elements with various levels of social richness on the web page. Cortese and Seo [39] manipulated the communication environment and forecasted the perceived social presence. Sia et al [40] confirmed that removing visual cues and providing anonymity reduces the perception of the level of social presence. However, relatively few existing studies have empirically investigated perceived social presence (PSP) in online health care and its effect on the mediated communication between physicians and patients.

Based on the literature review, we believe that social presence is an appropriate theoretical framework to investigate how online medical interactions with different levels of information richness influence patients' WTC.

Hypothesis Development and Research Model

Willingness to Communicate About Health

Many studies have discussed the effect of social presence on WTC online especially in the field of online learning [41,42] and revealed that students who perceive a higher level of social presence are usually more willing to communicate [43-45]. Although there is research showing that patients' WTC may help the treatment to be more effective [46], few studies have discussed patients' WTC in online medical services. Some scholars are interested in the relationship between social presence and WTC in the context of psychotherapy. For example, Cukor et al [47] found that in online medical services, videophones create a certain level of social presence in mental disease counseling and enable the patient to be more willing to discuss complex topics. A similar effect is expected to be found in online health care based on the theory of social presence. According to studies on the relationship between PSP and users' willingness in other fields, we suggest the following hypothesis:

H1. Participants' PSP is positively correlated with WTCH in online medical consultation.

Self-perceived Communication Competence

Communication competence is generally defined as the "cyclical process that leads to the continual refinement of one's social communication repertoire" [48], and it is a way of dynamic understanding instead of a personality trait [49]. Since it is difficult to conceptualize or measure communication competence, most behavior-oriented studies focus on "individuals' perceptions of competent communication behaviors" [48].

McCroskey and McCroskey [50] define self-perceived communication competence (SPCC) as an individual's evaluation of their confidence in their ability to communicate. It is the individual's self-perception instead of their actual competence or skill that initiates WTC [51]. The effect of communication competence has been verified in the medical field [52,53]. The Medical Communication Competence Scales developed by Cegala et al [54] help to further study the physician-patient interaction in medical interviews. In the CMC environment, patients' competence in medical communication will help them describe the physical condition to the physician [55].

Communication competence is seldom examined with social presence. Wrench and Punyanunt-Carter [56] demonstrated that there is a positive correlation between CMC competence and CMC presence. Although CMC competence is not equivalent to communication competence, it has been proven that CMC competence embodies the notion of communication competence. In addition, SPCC has been proven to have significant influence on people's WTC [57], and an individual's perception of their competence can overcome the influence of their actual competence on the WTC [58]. When an individual is motivated to communicate competently, their perception of competence increases [59] and their WTC will grow accordingly.

Therefore, we hypothesize the following:

H2. Participants' SPCC mediates the relationship between PSP and WTCH in online medical consultation.

Communication Apprehension

Studies on CA started from the 1970s. It refers to "an individual's level of fear or anxiety associated with either real or anticipated communication with another person or persons" [60]. In general, CA is a large topic and has various definitions. It is also conceptualized as (1) the personality trait an individual has in communication, which is enduring in communicative situations, or (2) the state of experience, which is situation-specific [61,62]. There are fewer people with trait CA than people with state CA [63].

Studies have proven the mutual influence between CA and social presence [39]. Ayres and Hopf [64] found that the perception of a human being can help to reduce CA. On the other hand, people who have experienced a high level of apprehension tend to perceive a lower level of social presence [56,65].

Few studies connect the application of CMC with CA, yet Burke et al [66] found that the application of CMC is strongly connected with the computer-mediated apprehension. People with higher level of CA perceive a lower level of social presence [39] and are more inclined to engage in CMC discussions [66]. In telemedicine, the presence of credible medical service providers or family members will help to reduce patients' apprehension effectively [67]. In addition, CA has significant influence on individuals' behaviors and motivations [68], especially on their WTC [39,66,69], considering both the amount and quality of the communication [70]. People with high levels of CA will try to keep silent or talk as little as possible to avoid communication, while those with low levels of CA usually seek

XSL•FO
RenderX

opportunities to communicate [71]. It does not necessarily mean that people with high levels of CA will totally avoid CMC, yet compared with those with low levels of CA, their WTC online does not seem to be very strong. Thus, we hypothesize the following:

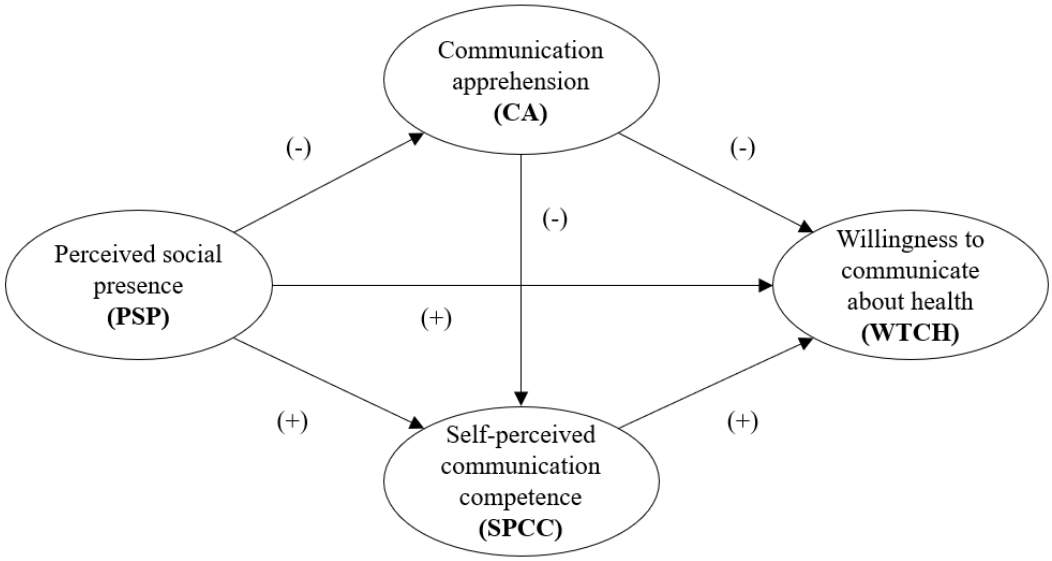
H3. Participants' CA mediates the relationship between PSP and WTCH in online medical consultation.

A high level of CA will lead to more negative evaluation of self-competence [72]. McCroskey et al [73] found that CA is negatively associated with SPCC, which has been further verified in other empirical studies [74,75]. People who are more apprehensive of communication tend to believe that they are less competent communicators [76,77]. Based on the findings of the above research studies, we hypothesize the following:

H4. Participants' CA and SPCC sequentially mediate the relationship between PSP and WTCH in online medical consultation.

Figure 1 presents the proposed research model which depicts both the direct relationship and the sequential indirect relationship between PSP and WTCH through CA and SPCC.

Figure 1. Research model.



Methods

Participants

We cooperated with a legally registered data company that helped us contact and recruit qualified participants living in Yunnan province via the internet with compensation. Only participants with their mobile IP located in Yunnan could take part in the experiment and complete the questionnaire. In total, 299 people participated in the experiment with 276 completing the whole process. However, 70 participants failed to pass the reverse coding test. Therefore, 206 valid *samples were finally collected (each compensated with 4 renminbi [RMB]), yielding a valid participation rate of 68.9% (see Table 1 for participant demographic information). According to a report by China Industrial Information [78], users aged 18 to 44 years account for the majority of all users on online medical platforms, with male users being 54.6% and female users being 45.4%. In our experiment, participants aged 18 to 44 years account for 88.8% (183/206), and male users and female users are 59.2% (122/206) and 40.8% (84/206), respectively, which is basically in line with the users on online medical platforms in China.

XSL•FO
RenderX

Table 1. Participant demographics.

Characteristics	Value, n (%)
Gender	
Female	84 (40.8)
Male	122 (59.2)
Age (years)	
<18	11 (5.3)
18-24	79 (38.4)
25-34	69 (33.5)
35-44	35 (17.0)
>44	12 (5.8)
Education level	
Primary school	1 (0.5)
Junior high school	25 (12.1)
High school	38 (18.5)
Junior college	58 (28.2)
Bachelor	75 (36.4)
Master	9 (4.4)
Doctor	0 (0)
Monthly income, RMB[a] (US$), n (%)	
<1000 (154)	15 (7.3)
1000-2000 (154-308)	7 (3.4)
2000-5000 (308-772)	61 (29.6)
5000-10,000 (772-1544)	102 (49.5)
10,000-100,000 (1544-15,440)	20 (9.7)
>100,000 (>15,440)	1 (0.5)

[a]RMB: renminbi.

Ethics Approval

The experiment was approved (H2021099I) by the Institutional Review Board for Human Research Protections of the Shanghai Jiao Tong University on March 23, 2021.

Materials

Our experiment used the stimuli investigated in our pretest. Group 1 was presented with a physician using a default profile photo in the system without any personal or professional information (Figure 2) while participants in group 2 were presented with a page illustrating the physician's profile including a photo, name, position, institution, expertise, etc, before the experiment started. Interactive experiences such as some random waiting time, a sign on the top indicating that the physician is entering the name and profile photo presented in the dialog box, etc, were also added in group 2 (Figure 3). Except for the above differences, the experiment was conducted under the same system settings for the 2 groups. In order to keep the consistency and effectiveness of the independent variables, the responses and diagnosis from the physician in the dialogs were based on the patient's selection of preset options.

We conducted a trial test on the stimuli of the experiment on November 11, 2020. We randomly invited 35 students (aged 17-23 years; male 57% [20/35], female 43% [15/35]) in a cafeteria on the campus of a major university in China to participate in our online experiment. The result of the independent samples *t* test shows that the levels of PSP are significantly different under 2 groups of stimuli (group 1: 2.68 [SD 0.39], group 2: 3.38 [SD 0.52]; *P*<.001).

XSL•FO
RenderX

Figure 2. Interface for group 1 with a low level of social presence.

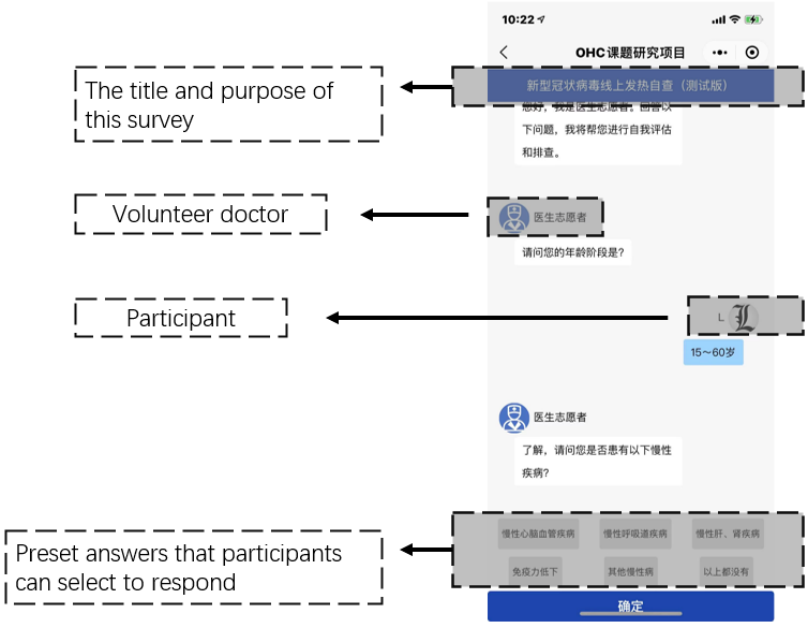
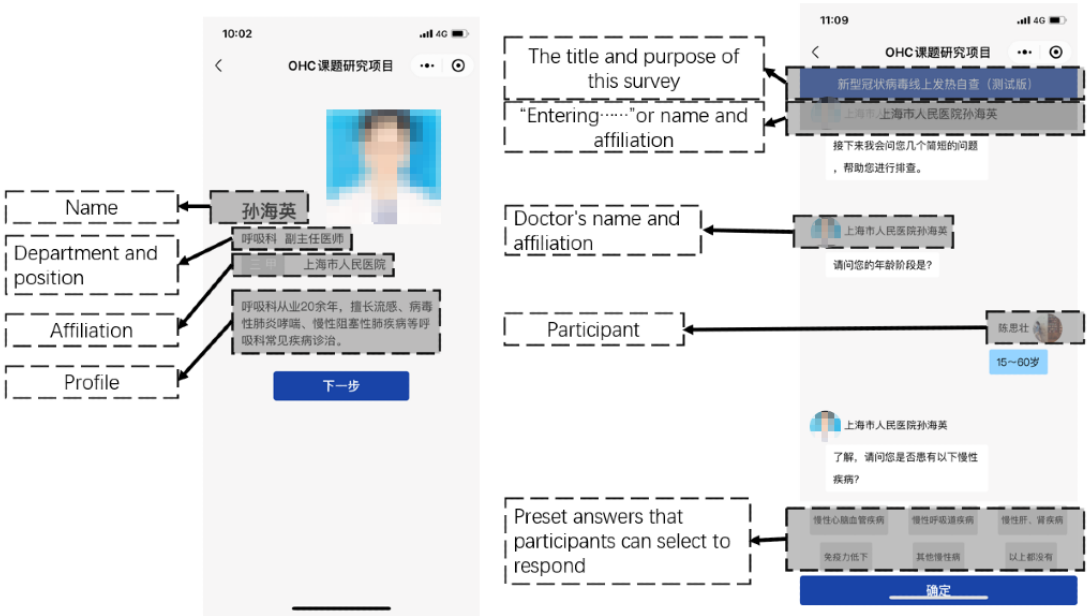


Figure 3. Interface for group 2 with a high level of social presence.



Design and Procedure

We have designed a 1-factorial experiment on 2 groups with the manipulation of social presence level. Participants were assigned randomly into 2 groups: group 1 with a low level of social presence and group 2 with a high level of social presence. Beside the stimuli, other system settings were the same for the 2 groups. Before the experiment started, participants were asked to provide demographic information and then, to improve the authenticity of the task, participants were invited to a pretest of an online system to self-check the potential risk of COVID-19 infection. Participants were informed that the test controller was expecting their honest feedback to further optimize the system, but the result of the test could not be regarded as clinical evidence. In the experiment, there was a reminder for the participants stating "If you feel sick, you need to visit a physician as soon as possible." On March 30, 2021, 6 new cases of COVID-19 were confirmed in Yunnan province, the first since September 2020 when the number of local confirmed cases fell to zero. Our experiment (H2021099I) ran from April 8-21, 2021, because some areas in Yunnan were rated as medium to high risk areas throughout the period, suitable for the experiment since people living in Yunnan might be concerned about the risk of infection. To simulate the mobile medical service environment, our experiment was conducted entirely on smartphones involving 2 groups of participants (106 in group 1 and 100 in group 2). The experiment was followed by a questionnaire.

XSL•FO
RenderX

Measurements

Social Presence

We adopted the 5-item PSP scale [30], which is frequently applied to examine the PSP in CMC. The scale ranges from 1 (strongly disagree) to 5 (strongly agree).

Communication Apprehension

We developed the 5-item Patients' Report of Communication Apprehension with Physicians proposed by Ayres et al [79] based on personal perceived communication apprehension. The scale measures patients' CA in physician-patient interactions on a scale from 1 (strongly disagree) to 5 (strongly agree).

Self-perceived Communication Competence

We developed the Patients' Self-competence Items Comprising the Medical Communication Competence Scale proposed by Cegala et al [54], ranging from 1 (strongly disagree) to 5 (strongly agree).

Willingness to Communicate About Health

We developed the WTCH scales proposed by Wright et al [80]. This scale has 10 items divided into 2 dimensions, provider and nonprovider. It is applied to measure whether a patient is willing to communicate about their health status with different people. Since the experiment was conducted entirely on smartphones, we used the 5 items from the provider dimension only, ranging from 1 (strongly disagree) to 5 (strongly agree).

Willingness to Communicate

We tried to control the effects of personality types and developed the WTC scales proposed by McCroskey [81]. In this scale, WTC includes 4 types of communicative situations and 3 kinds of audiences. Based on our experiment, we adopt the 4 items associated with the investigation of the subscores related to strangers. Participants indicated the probability they would choose to communicate by reporting the percentage from 0 (never) to 100 (always).

Validity and Reliability Check

The 4 constructs (PSP, CA, SPCC, and WTCH) in the questionnaire were all developed from scales in prior studies, and the content validity of these constructs has been repeatedly verified (see Multimedia Appendix 1). In addition, in our pretest we invited 35 participants on campus to give us some advice on the understanding, wording, and readability of the items after they completed the experiment and the questionnaire. Their suggestions were fully considered and helped us to reach the final version of the questionnaire (Multimedia Appendix 2).

Confirmatory factor analysis was run in SPSS AMOS (version 23.0, IBM Corp) to check the convergent validity and reliability of the construct. Confirmatory factor analysis fit information indicated that the measurement model was acceptable (χ^2/df=1.93, root mean square error of approximation=0.07, comparative fit index=0.92, Tucker-Lewis index=0.90, incremental fit index=0.92; Table 2). According to Fornell and Larcker [82], when the average variance extracted (AVE) of the construct is equal or greater than 0.50 and its construct reliability (CR) is greater than 0.70, the construct passes the convergent validity test. The standard load factors of the items in construct PSP ranged from 0.72 to 0.82 (P<.001, AVE=0.57, CR=0.87). The factor of CA ranged from 0.58 to 0.84 (P<.001, AVE=0.52, CR=0.81). Since the 6 items in construct SPCC are comprised of 2 dimensions (items 1 to 4 from 1 dimension and items 5 and 6 from the other), they were analyzed in 2 separate factors. The standard load factors of items 1 to 4 ranged from 0.70 to 0.94 (P<.001, AVE=0.62, CR=0.76), while the factors of items 5 and 6 ranged from 0.76 to 0.81 (P<.001, AVE=0.62, CR=0.89). The standard load factor of the items in construct WTCH ranged from 0.66 to 0.80 (P<.001, AVE=0.52, CR=0.84). The standard load factor of the items in construct WTC ranged from 0.54 to 0.87 (P<.001, AVE=0.51, CR=0.80). Therefore, the convergent validity and CR of our questionnaire meets the requirement.

XSL•FO
RenderX

Table 2. Results of confirmatory factor analysis.

Factor and item	Std[a] loading	Convergent validity	
		CR[b]	AVE[c]
PSP[d]	—[e]	0.87	0.57
PSP1	0.74	—	—
PSP2	0.72	—	—
PSP3	0.76	—	—
PSP4	0.82	—	—
PSP5	0.74	—	—
CA[f]	—	0.81	0.52
CA1	0.79	—	—
CA2	0.84	—	—
CA3	0.58	—	—
CA4	0.63	—	—
SPCC[g] (1-4)	—	0.89	0.68
SPCC1	0.86	—	—
SPCC2	0.94	—	—
SPCC3	0.70	—	—
SPCC4	0.77	—	—
SPCC (5-6)[h]	—	0.76	0.62
SPCC5	0.81	—	—
SPCC6	0.76	—	—
WTCH[i]	—	0.84	0.52
WTCH1	0.69	—	—
WTCH2	0.69	—	—
WTCH3	0.80	—	—
WTCH4	0.66	—	—
WTCH5	0.74	—	—
WTC[j]	—	0.80	0.51
WTC1	0.81	—	—
WTC2	0.58	—	—
WTC3	0.54	—	—
WTC4	0.87	—	—

[a]Std: standardized.
[b]CR: construct reliability.
[c]AVE: average variance extracted.
[d]PSP: perceived social presence.
[e]Not applicable.
[f]CA: communication apprehension.
[g]SPCC: self-perceived communication competence dimension 1.
[h]SPCC (5-6): self-perceived communication competence dimension 2.
[i]WTCH: willingness to communicate about health.
[j]WTC: willingness to communicate.

XSL•FO
RenderX

Results

Manipulation Check

The interfaces of our online experiment were designed to present 2 levels (high and low) of social presence. The independent samples *t* test was applied to examine whether the manipulation successfully triggered different levels of social presence as perceived by participants. The test proved that the PSP of group 2 was significantly higher than that of group 1 (P=.04, F=0.01, t=–2.12). Table 3 shows the results of the *t* test.

Table 3. Results of independent samples *t* test.

Group	Social presence level	Mean (SD)	*F* score	*t* score	*P* value
1	Low	3.85 (0.77)	0.01	–2.12	.04
2	High	4.07 (0.75)	—[a]	—	—

[a]Not applicable.

Hypothesis Testing

Based on the proposed model of this study, model 6 described in the bootstrap methods of SPSS PROCESS developed by Hayes [83] was applied to analyze the sequential mediating effect of the model. In this study, the average score of the items in the construct was calculated to represent the value of each construct. In the model, PSP is the independent variable (X), WTCH is the dependent variable (Y), CA is mediator 1 (M_1), and SPCC is mediator 2 (M_2).

Correlation Test

Based on the proposed model of this study, a partial correlation test controlling the effects of age, gender, education level, income level, and WTC was conducted to examine the intercorrelations among the 4 variables. Results in Table 4 show that the intercorrelations among all the variables are significant (P<.001) and the signs of the coefficients are consistent with the model prediction.

The total, direct, and indirect effects between PSP and WTCH with the mediation of CA and SPCC are calculated by model 6, controlling the effects of age, gender, education level, income level, and WTC.

Table 4. Partial correlation analysis controlling age, gender, education level, income level, and willingness to communicate.

	PSP[a]	CA[b]	SPCC[c]	WTCH[d]
PSP	—[e]	—	—	—
CA	–.23	—	—	—
SPCC	0.37	–.40	—	—
WTCH	0.64	–.35	0.6	—

[a]PSP: perceived social presence.

[b]CA: communication apprehension.

[c]SPCC: self-perceived communication competence.

[d]WTCH: willingness to communicate about health.

[e]Not applicable.

Total Effect

The regression effect between PSP (X) and WTCH (Y) is significant. PSP positively affects WTCH, supporting H1. The total effect of PSP on WTCH is 0.56 (P<.001, t=11.63, R^2=0.42, F=23.71).

Direct Effect

According to the results calculated based on model 6, the direct effect of PSP on WTCH accounts for 74.37% (P<.001, effect=0.41, t=9.30) of the total effect, indicating that 25.63% of the total effect is indirect.

Indirect Effect

The results in Tables 5 and 6 show that the coefficient of PSP→CA, PSP→SPCC, CA→SPCC, PSP→WTCH, and SPCC→WTCH are all significant while the coefficient of CA→WTCH is not significant, indicating that the indirect pathway of PSP→SPCC→WTCH is supported (supporting H2) and the sequential mediation of the pathway of PSP→CA→SPCC→WTCH is supported (supporting H4), but the indirect pathway of PSP→CA→WTCH is not supported (rejecting H3). It indicates that CA alone cannot mediate the relationship between PSP and WTCH. Instead, CA affects WTCH only through the mediating effect of SPCC.

Table 7 summarizes the indirect effects of the 3 pathways. According to Hayes, the pathway is significant when the BootLLCI and BootULCI are both above zero or below zero [83]. Therefore, the pathway of PSP→CA→WTCH is not significant while the other 2 are significant, repeating the results from the equations. The effect of the pathway of PSP→SPCC→WTCH accounts for 68.75% of the total indirect effect and for the pathways of PSP→CA→SPCC→WTCH and PSP→CA→WTCH, the ratios are 19.24% and 12.01%,

XSL•FO
RenderX

respectively. It can be inferred that the effect through the mediator SPCC accounts for 87.99% of the total indirect effect, indicating that SPCC plays a more important role in this model.

Based on the coefficients of the causal paths illustrated in Tables 5 and 6, the final model is presented in Figure 4.

Table 5. Summary of the coefficients and the models controlling age, gender, education level, income level, and willingness to communicate.

	CA (M_1)[a]			SPCC (M_2)[b]			WTCH (Y)[c]		
	Coeff.	*t*	*P* value	Coeff.	*t*	*P* value	Coeff.	*t*	*P* value
PSP (X)[d]	–.26	–3.38	<.001	0.21	4.54	<.001	0.41	9.3	<.001
CA (M_1)	—	—	—	–.24	–5.14	<.001	–.08	–1.75	.082
SPCC (M_2)	—	—	—	—	—	—	0.47	7.2	<.001

[a]CA (M_1): communication apprehension as mediator 1.

[b]SPCC (M_2): self-perceived communication competence as mediator 2.

[c]WTCH (Y): willingness to communicate about health as Y.

[d]PSP (X): perceived social presence as X.

Table 6. Model summary of communication apprehension, self-perceived communication competence, and willingness to communicate about health as outcome variables.

	CA (M_1)[a]	SPCC (M_2)[b]	WTCH (Y)[c]
R^2	0.11	0.29	0.57
F score	3.92	11.48	33.05
P value	.001	<.001	<.001

[a]CA (M_1): communication apprehension as mediator 1.

[b]SPCC (M_2): self-perceived communication competence as mediator 2.

[c]WTCH (Y): willingness to communicate about health as Y.

Table 7. Summary of the mediation pathways.

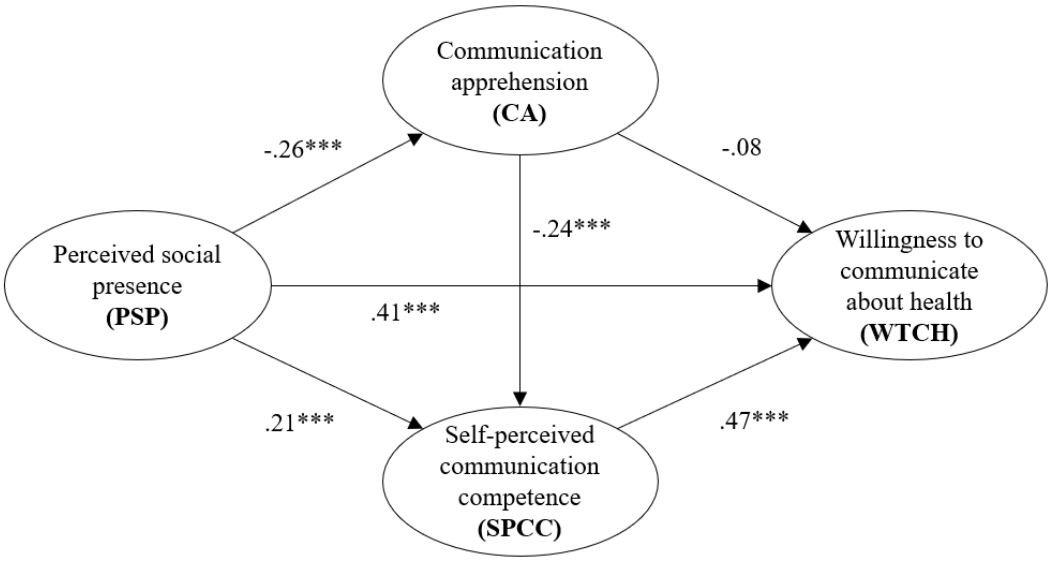
Mediation pathways	Effect	BootLLCI[a]	BootULCI[b]
Total	0.14	0.087	0.208
PSP→CA→WTCH	0.02	–0.001	0.048
PSP→SPCC→WTCH	0.10	0.054	0.150
PSP→CA→SPCC→WTCH	0.03	0.009	0.052

[a]BootLLCI: bootstrap lower limit confidence interval.

[b]BootULCI: bootstrap upper limit confidence interval.

XSL•FO
RenderX

Figure 4. Final mediation model. ***P<.001.



Discussion

Principal Findings

Prior research models consider that PSP is influenced by CA in computer-mediated environments; to be specific, people with a higher level of CA will experience a lower level of PSP [39,56]. However, this is the opposite of the causality shown by our experiment results. Our research finds that the level of PSP will negatively influence people's CA, yet the causality of this influence has been proven by Burke et al [66] to be the opposite—people experiencing a high level of CA tend to be involved in CMC discussions with a low level of social presence. In addition, Burke et al [66] found that PSP is negatively associated with WTCH, which is also the opposite of our results that clearly show a positive relationship between PSP and people's communication willingness mediated by computers.

Prior literature has demonstrated the correlation between SPCC and WTCH [84-86], and such correlation is also applicable in the context of online medical communication studied in this research. Further findings show that SPCC mediates the relationship between PSP and WTCH. As an individual perception, SPCC is a key factor affecting users' WTC in the CMC environment. It seems that SPCC is less associated with the individuals' innate and actual communication skills but is more dependent on their perception of their abilities. And such perception will directly influence users' WTC by increasing or decreasing social cues in the communication environment.

Beside the mediating effects of CA and SPCC discussed in this study, we found that the effect of CA on WTCH is exerted entirely through the mediating effect of SPCC. The perspective that health communication competence is one of the best predictors of patients' WTC [87] has been proven again. Competence is a sociopsychological concept. It is more like a response to a specific situation or social context to achieve the communicative goals [88]. McCroskey [51] believes that whether people decide to start a conversation usually depends on self-perceived rather than actual communicative competence. People experiencing a high level of CA will decrease their evaluation of their communicative competence [72]. The sequential mediating effect of CA and SPCC found in this study also proves that in the environment of online mobile medical services, CA cannot hinder or facilitate communication directly—instead, it should transfer its effect to SPCC. If participants experience a higher level of CA in the experiment, their SPCC will be decreased accordingly, thus showing a lower WTC.

Comparison With Prior Work

Although prior studies have examined and confirmed the influence of social presence in the medical field [34,35], due to the lack of an accepted definition and difficulty of measurement, few empirical studies have explored the relationship between social presence and WTC, let alone the exploration of the causal relationship between them. Thus, our study has supplemented current research results to a certain extent by conducting the experiment on a more diverse group of people.

In prior studies, the interrelationships among CA, SPCC, and WTC have received much attention. However, the difference in the communication performance of people with different levels of CA, especially the difference in WTC under the influence of online and offline CA and the different levels of CA when facing different people, have not been fully discussed. Some studies hold that face-to-face CA seems to motivate people to communicate online [89]. In the CMC environment which is lower in media richness, what is the influence of the state CA on WTC? In addition, the offline SPCC is decisive for the individual's WTC. Will the online SPCC in some specific situations vary with the increase or decrease in social cues? To answer these questions, we simulates the online mobile medical environment and manipulates participants' PSP by stimulating their imagination of interaction and providing actual interactions with the others, which supplements existing studies of the antecedents of WTC in the computer-mediated environment. Although our results show that the mediating effect of CA on the relationship between PSP and WTCH is not supported, there are some valuable findings in our research.

XSL•FO
RenderX

Practical Implications

Our study will provide some practical inspiration for the popularization of online medical service in the future, especially for the promotion of online physician-patient communication. To construct a friendly environment for medical communication, the program designer should fully take users' PSP into consideration. For online patients, if they have worries or lack confidence in the CMC with physicians, their SPCC will be influenced and their WTCH conditions will be affected accordingly. Given the important role of SPCC, online medical service providers should not only help users to acquire substantial communication skills, but also build a friendly consultation environment. Interactive elements enabling users to feel more confident can be applied to optimize the interface (eg, a typing prompt function to suggest expressions users may proceed with in the conversation). When users feel more confident about their competence to initiate and continue an effective communication, they are more likely to be engaged in online consultation. To take care of users' CA, which is an influencing factor of SPCC, online health service providers should try to ease users' stress and enhance their sense of security by scientific communication skills catering for users' emotions. Experiencing a lower level of CA will result in an increase in SPCC, which will encourage users to engage in online health communication.

Limitations

First, concerning our research model, the experiment only investigates the users' WTC, and the amount of actual communication (or expressions) has not been measured. In fact, despite their unwillingness, people with a high level of CA may not have poor performance in communication [70]. Second, our experiment only investigates the text communication between physicians and patients online. Users' WTC online through audio or video is also worth studying.

Concerning the design of the experimental stimuli, both the self-examination of the risk of COVID-19 infection and the time of conducting the experiment are cross-sectional and have their inherent limitations. If the participant wants to have medical services online and interact with a real physician, their communication will adjust to the feedbacks from each other. However, in order to control the manipulation of PSP, we have eliminated the personalized feedback from the simulated online physician. Our experiment is based on the Chinese cultural context and medical conditions. If the experiment is conducted with a more international and diverse group of participants, can the findings of this study be repeated? More proof is needed.

Finally, our study focuses more on patients; however, it has been proven that the physician-patient communication is more physician-centric in online medical services than in face-to-face consultations [46]. If researchers are interested in the field of online physician-patient communication, especially the patient's WTC, we suggest they explore more on physician-centric studies.

Conclusions

The computer-mediated environment was once considered as an inactive medium for the communication of health problems. Responding to the theory of social presence, we have verified the pathway of social presence's influence on patients' WTC. Specifically, this study examined the indirect effects of the 2 parallel mediators. The results of the experiment show that although only the indirect path via SPCC is supported, some innovative findings can still be drawn. Considering the diverse definitions of social presence, a convincing measurement remains to be developed and tested. Our experiment has proved that the increment of social cues will positively influence the degree of social presence perceived by participants in mobile medical consultations. When provided with more personal information on the physician and more interactivity, participants will perceive the social presence to be more prominent. Otherwise, there will be subtle perception of social presence. Given the identification of a reliable manipulation that can be applied to measure social presence as the embodiment of the authenticity of others in communication, future research studies can focus more on evaluating the antecedents and consequences of social presence. For direct effect, the results demonstrate that patients' WTCH is directly influenced by PSP. For indirect effect, patients' WTC will increase by the higher level of their SPCC. However, a lower level of CA will not lead to an increased level of WTC, which is inconsistent with prior studies and contributes to the literature of social presence theory in the field of health communication. In general, this study developed and explored the concept and analysis framework of PSP that affects patients' WTC with physicians about their health conditions in the context of mobile health care and verified the mediating effect of online experiential interaction between physicians and patients on the latter's WTC. The research on the impact of PSP on patients' willingness to talk about health online is only a preliminary attempt. The process and outcome of online physician-patient communication are affected by multiple factors that need to be explored more fully in the future.

Authors' Contributions

LC conceptualized the theoretical framework and built the model. LC and DZ collected and analyzed the data and wrote the first draft of the manuscript. MH revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Construct items and sources.
[DOC File , 46 KB - jmir_v24i5e31797_app1.doc]

Multimedia Appendix 2
Questionnaire.
[DOCX File , 25 KB - jmir_v24i5e31797_app2.docx]

References

1. Lu X, Zhang R. Impact of physician-patient communication in online health communities on patient compliance: cross-sectional questionnaire study. J Med Internet Res 2019 May 13;21(5):e12891 [FREE Full text] [doi: 10.2196/12891] [Medline: 31094342]
2. Chen X. Research on the characteristics and communication mode of virtual community: take DXY community as an example. In: Proceedings of CCNT 2018. 2018 Presented at: 2018 International Conference on Computer, Communication and Network Technology; June 29-30; Wuzhen. [doi: 10.12783/dtcse/ccnt2018/24781]
3. Wilson TD. Human information behavior. Informing Sci J 2000;3:49-56. [doi: 10.28945/576]
4. McCroskey J, Baer J. Willingness to communicate: the construct and its measurement. 1985 Presented at: 71st Annual Meeting of the Speech Communication Association; 1985; Denver URL: https://files.eric.ed.gov/fulltext/ED265604.pdf
5. Shahbaz M, Khan MS, Ishtiaq Khan RM, Mustafa G. Role of self-perceived communication competence and communication apprehension for willingness to communicate in L1 and L2. J Educ Soc Res 2016 Jan 5;6(1):1. [doi: 10.5901/jesr.2016.v6n1p158]
6. Xiao S. Tension in the Chinese Doctor-Patient-Family Relationship: A Qualitative Study in Hunan Province, China [Dissertation]. New Haven: Yale University; 2018.
7. Wright KB, Frey LR. Communication and care in an acute cancer center: the effects of patients' willingness to communicate about health, health-care environment perceptions, and health status on information seeking, participation in care practices, and satisfaction. Health Commun 2008 Jul 22;23(4):369-379. [doi: 10.1080/10410230802229886] [Medline: 18702001]
8. Jiang S. The relationship between face-to-face and online patient-provider communication: examining the moderating roles of patient trust and patient satisfaction. Health Commun 2020 Mar 01;35(3):341-349. [doi: 10.1080/10410236.2018.1563030] [Medline: 30821506]
9. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
10. Parsons T. The Social System. East Sussex: Psychology Press; 1991.
11. Cline R. Everyday interpersonal communication and health. In: Thompson T, Dorsey A, Miller K, editors. Handbook of Health Communication. Mahwah: Lawrence Erlbaum Associates Publishers; 2003:285-313.
12. Smith SW, Kopfman JE, Lindsey LLM, Yoo J, Morrison K. Encouraging family discussion on the decision to donate organs: the role of the willingness to communicate scale. Health Commun 2004 Jul;16(3):333-346. [doi: 10.1207/s15327027hc1603_4]
13. McComas KA, Yang Z, Gay GK, Leonard JP, Dannenberg AJ, Dillon H. Individuals' willingness to talk to their doctors about clinical trial enrollment. J Health Commun 2010 Mar;15(2):189-204. [doi: 10.1080/10810730903528058] [Medline: 20390986]
14. Canzona MR, Fisher CL, Wright KB, Ledford CJW. Talking about sexual health during survivorship: understanding what shapes breast cancer survivors' willingness to communicate with providers. J Cancer Surviv 2019 Dec 18;13(6):932-942. [doi: 10.1007/s11764-019-00809-2] [Medline: 31741248]
15. Petrič G, Atanasova S, Kamin T. Impact of social processes in online health communities on patient empowerment in relationship with the physician: emergence of functional and dysfunctional empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
16. Roettl J, Bidmon S, Terlutter R. What predicts patients' willingness to undergo online treatment and pay for online treatment? Results from a web-based survey to investigate the changing patient-physician relationship. J Med Internet Res 2016 Feb 04;18(2):e32 [FREE Full text] [doi: 10.2196/jmir.5244] [Medline: 26846162]
17. Wright KB, Frey L, Sopory P. Willingness to communicate about health as an underlying trait of patient self-advocacy: the development of the willingness to communicate about health (WTCH) measure. Commun Stud 2007 Feb 16;58(1):35-51. [doi: 10.1080/10510970601168673]
18. Pronk NP, Tan AW, O'Connor P. Obesity, fitness, willingness to communicate and health care costs. Med Sci Sports Exerc 1999 Nov;31(11):1535-1543. [doi: 10.1097/00005768-199911000-00007] [Medline: 10589854]
19. Baker SC, Watson BM. Investigating the association between internet health information use and patient willingness to communicate with health care providers. Health Commun 2020 May 18;35(6):716-725. [doi: 10.1080/10410236.2019.1584778] [Medline: 30880472]

XSL•FO
RenderX

20. Parker EB, Short J, Williams E, Christie B. The social psychology of telecommunications. Contemp Sociol 1978 Jan;7(1):32. [doi: 10.2307/2065899]
21. Fulk J, Steinfield CW, Schmitz J, Power JG. A social information processing model of media use in organizations. Commun Res 2016 Jun 30;14(5):529-552. [doi: 10.1177/009365087014005005]
22. Biocca F, Harms C, Burgoon JK. Toward a more robust theory and measure of social presence: review and suggested criteria. Presence Teleoperators Virt Environ 2003 Oct;12(5):456-480. [doi: 10.1162/105474603322761270]
23. Nowak KL, Biocca F. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence Teleoperators Virtual Environ 2003 Oct;12(5):481-494. [doi: 10.1162/105474603322761289]
24. Yamada M. The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: experimental study. Comput Educ 2009 May;52(4):820-833. [doi: 10.1016/j.compedu.2008.12.007]
25. Garrison D, Anderson T, Archer W. A theory of critical inquiry in online distance education. In: Handbook of Distance Education. Mahwah: Lawrence Erlbaum Associates Publishers; 2003:113-127.
26. Fägersten K. Using discourse analysis to assess social co-presence in the video conference environment. In: Shedletsky L, Aitken J, editors. Cases on Online Discussion and Interaction: Experiences and Outcomes. Hershey: IGI Global; 2010:175-193.
27. Yoo Y, Alavi M. Media and group cohesion: relative influences on social presence, task participation, and group consensus. Manag Inf Syst Q 2001 Sep;25(3):371-390. [doi: 10.2307/3250922]
28. Siegel J, Dubrovsky V, Kiesler S, McGuire TW. Group processes in computer-mediated communication. Organ Behav Hum Decis Processes 1986 Apr;37(2):157-187. [doi: 10.1016/0749-5978(86)90050-6]
29. Hancock JT, Dunham PJ. Impression formation in computer-mediated communication revisited. Commun Res 2016 Jun 29;28(3):325-347. [doi: 10.1177/009365001028003004]
30. Gefen D, Straub DW. Managing user trust in B2C e-services. e-Serv J 2003;2(2):7-24. [doi: 10.1353/esj.2003.0011]
31. Gefen D, Straub DW. Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. Omega 2004 Dec;32(6):407-424. [doi: 10.1016/j.omega.2004.01.006]
32. Cyr D, Hassanein K, Head M, Ivanov A. The role of social presence in establishing loyalty in e-Service environments. Interact Comput 2007 Jan;19(1):43-56. [doi: 10.1016/j.intcom.2006.07.010]
33. Mäntymäki M, Salo J. Trust, social presence and customer loyalty in social virtual worlds. 2010 Presented at: Bled Proceedings; 2010; Bled URL: https://aisel.aisnet.org/bled2010/38
34. Goonawardene N, Tan S. Practitioner-driven virtual communities: an attachment theory perspective to patients' adherence to online health advice. 2014. URL: https://aisel.aisnet.org/icis2014/proceedings/ISHealthcare/31/ [accessed 2022-04-21]
35. Peng X, Sun D, Zhao Y, Xu W. What trigger people use physician-patient interactive OHCs? An empirical research based integration model. PACIS 2015 Proceedings 2. 2015. URL: http://aisel.aisnet.org/pacis2015/2 [accessed 2022-04-21]
36. Rourke L, Anderson T, Garrison D, Archer W. Assessing social presence in asynchronous text-based computer conferencing. J Dist Educ 1999;14(2):50-71 [FREE Full text]
37. Straub D, Karahanna E. Knowledge worker communications and recipient availability: toward a task closure explanation of media choice. Organ Sci 1998 Apr;9(2):160-175. [doi: 10.1287/orsc.9.2.160]
38. Hassanein K, Head M. Manipulating perceived social presence through the web interface and its impact on attitude towards online shopping. Int J Hum Comput Stud 2007 Aug;65(8):689-708. [doi: 10.1016/j.ijhcs.2006.11.018]
39. Cortese J, Seo M. The role of social presence in opinion expression during FtF and CMC discussions. Commun Res Rep 2012 Jan;29(1):44-53. [doi: 10.1080/08824096.2011.639913]
40. Sia C, Tan BCY, Wei K. Group polarization and computer-mediated communication: effects of communication cues, social presence, and anonymity. Inf Syst Res 2002 Mar;13(1):70-90. [doi: 10.1287/isre.13.1.70.92]
41. Le T, Cunningham U, Watson K. The relationship between willingness to communicate and social presence in an online English language course. Jalt Call J 2018;14(1):43-59 [FREE Full text] [doi: 10.29140/jaltcall.v14n1.223]
42. Al-Amrani S, Harrington M. The impact of online social presence on Omani female students' willingness to communicate in English. CALL-EJ. 2: 21; 2020. URL: http://callej.org/journal/21-2/Al-Amrani-Harrington2020.pdf [accessed 2022-04-21]
43. Tu C, McIsaac M. The relationship of social presence and interaction in online classes. Am J Dist Educ 2002 Sep;16(3):131-150. [doi: 10.1207/s15389286ajde1603_2]
44. Garrison DR, Cleveland-Innes M. Facilitating cognitive presence in online learning: interaction is not enough. Am J Dist Educ 2005 Sep;19(3):133-148. [doi: 10.1207/s15389286ajde1903_2]
45. Swan K, Shih L. On the nature and development of social presence in online course discussions. Online Learn J. 2005. URL: https://olj.onlinelearningconsortium.org/index.php/olj/article/view/1788 [accessed 2022-04-21]
46. Lu X, Zhang R. Impact of patient information behaviours in online health communities on patient compliance and the mediating role of patients' perceived empathy. Patient Educ Couns 2021 Jan;104(1):186-193. [doi: 10.1016/j.pec.2020.07.001] [Medline: 32665071]

XSL•FO
RenderX

47. Cukor P, Baer L, Willis BS, Leahy L, O'Laughlen J, Murphy M, et al. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. Telemed J 1998 Jan;4(4):313-321. [doi: 10.1089/tmj.1.1998.4.313] [Medline: 10220471]
48. Jablin F, Sias P. Communication competence. In: Jablin F, Putnam L, editors. The New Handbook of Organizational Communication: Advances in Theory, Research, and Methods. Thousand Oaks: Sage Publications; 2001:819-864.
49. Parks M. Communicative competence and interpersonal control. In: Knapp M, Daly L, editors. Handbook of Interpersonal Communication 3rd Edition. Thousand Oaks: Sage Publications; 2001:589-618.
50. McCroskey JC, McCroskey LL. Self‐report as an approach to measuring communication competence. Commun Res Rep 1988 Dec;5(2):108-113. [doi: 10.1080/08824098809359810]
51. McCroskey J. Willingness to communicate, communication apprehension, and self-perceived communication competence: conceptualizations and perspectives. In: Daly J, McCroskey J, Ayres J, Hopf T, Ayres M, editors. Avoiding Communication: Shyness, Reticence, and Communication Apprehension 2nd Edition. Cresskill: Hampton Press Inc; 1997:75-108.
52. McGee DS, Cegala DJ. Patient communication skills raining for improved communication competence in the primary care medical consultation. J Appl Commun Res 1998 Nov;26(4):412-430. [doi: 10.1080/00909889809365517]
53. Bylund C, Brown R, Lubrano di Ciccone B, Diamond C, Eddington J, Kissane D. Assessing facilitator competence in a comprehensive communication skills training programme. Med Educ 2009 Apr;43(4):342-349. [doi: 10.1111/j.1365-2923.2009.03302.x] [Medline: 19335576]
54. Cegala DJ, Coleman MT, Turner JW. The development and partial assessment of the medical communication competence scale. Health Commun 1998 Jul;10(3):261-288. [doi: 10.1207/s15327027hc1003_5] [Medline: 16370986]
55. Avtgis TA, Brann M, Staggers SM. Perceived information exchange and health control expectancies as influenced by a patient's medical interview situation. Commun Res Rep 2006 Dec;23(4):231-237. [doi: 10.1080/08824090600962284]
56. Wrench JS, Punyanunt-Carter NM. The relationship between computer-mediated-communication competence, apprehension, self-efficacy, perceived confidence, and social presence. South Commun J 2007 Oct 26;72(4):355-378. [doi: 10.1080/10417940701667696]
57. Aliakbari M, Kamangar M, Khany R. Willingness to communicate in English among Iranian EFL students. Engl Lang Teach 2016 Apr 05;9(5):33-45. [doi: 10.5539/elt.v9n5p33]
58. McCroskey JC, Richmond VP. Willingness to communicate: differing cultural perspectives. South Commun J 1990 Dec;56(1):72-77. [doi: 10.1080/10417949009372817]
59. Morreale S, Spitzberg B, Barge J. Communication: Motivation, Knowledge, Skills 3rd edition. New York: Peter Lang Inc; 2013.
60. McCROSKEY JC. Oral communication apprehension: a summary of recent theory and research. Human Comm Res 1977 Sep;4(1):78-96. [doi: 10.1111/j.1468-2958.1977.tb00599.x]
61. McCroskey JC, Richmond VP. Communication apprehension and shyness: conceptual and operational distinctions. Cent States Speech J 2009 May 22;33(3):458-468. [doi: 10.1080/10510978209388452]
62. McCroskey J, Beatty M. Oral communication apprehension. In: Jones W, Cheek J, Briggs S, editors. Shyness: Perspectives on Research and Treatment (Emotions, Personality, and Psychotherapy). New York: Springer; 1986:279-293.
63. Miller L. Communication Apprehension and Its Effects on Online Readiness Skills [Thesis]. Spokane: Gonzaga University; 2014.
64. Ayres J, Hopf TS. The long‐term effect of visualization in the classroom: a brief research report. Commun Educ 2009 May 18;39(1):75-78. [doi: 10.1080/03634529009378788]
65. Tu C. On-line learning migration: from social learning theory to social presence theory in a CMC environment. J Netw Comput Appl 2000 Jan;23(1):27-37. [doi: 10.1006/jnca.1999.0099]
66. Burke B, Cameron K, Cooper J, Johnson S, Miller K. The correlation between computer mediated communication and communication apprehension. 2014. URL: https://journals.chapman.edu/ojs/index.php/mc/article/view/897/1044 [accessed 2022-04-21]
67. Kavanagh SJ, Yellowlees PM. Telemedicine: clinical applications in mental health. Aust Fam Physician 1995 Jul;24(7):1242-1247. [Medline: 7661778]
68. Flaherty LM, Pearce KJ, Rubin RB. Internet and face-to-face communication: not functional alternatives. Commun Q 2009 May 21;46(3):250-268. [doi: 10.1080/01463379809370100]
69. Limpornpugdee W, Janz B, Richardson S. Communication competence factors as moderators to the relationship between user participation and information quality. J Inf Technol Manag 2009;20(4):1-21 [FREE Full text]
70. Allen M, Bourhis J. The relationship of communication apprehension to communication behavior: a meta‐analysis. Commun Q 2009 May 21;44(2):214-226. [doi: 10.1080/01463379609370011]
71. Mccroskey JC. Oral communication apprehension: a reconceptualization. Ann Int Commun Assoc 2016 May 18;6(1):136-170. [doi: 10.1080/23808985.1982.11678497]
72. Bandura A. Self-efficacy conception of anxiety. Anxiety Res 2007 May 29;1(2):77-98. [doi: 10.1080/10615808808248222]
73. McCroskey JC, Burroughs NF, Daun A, Richmond VP. Correlates of quietness: Swedish and American perspectives. Commun Q 2009 May 21;38(2):127-137. [doi: 10.1080/01463379009369749]

XSL•FO
RenderX

74. Blood GW, Blood IM, Tellis G, Gabel R. Communication apprehension and self-perceived communication competence in adolescents who stutter. J Fluency Disord 2001 Sep;26(3):161-178. [doi: 10.1016/s0094-730x(01)00097-3]
75. Subekti AS. Self-perceived communication competence and communication apprehension: a study of Indonesian college students. J Engl Educ Lit Cult 2020 Feb 29;5(1):4-31. [doi: 10.30659/e.5.1.14-31]
76. Macintyre P, Noels K, Clément R. Biases in self-ratings of second language proficiency: the role of language anxiety. Lang Learn 1997;47(2):265-287. [doi: 10.1111/0023-8333.81997008]
77. Tamara S, Setiyadi A, Nainggolan F. The role of self-perceived communication competence, communication apprehension, and motivation towards willingness to communicate of the second year science class students of SMAN 9 Bandar Lampung. Unila J Engl Teach 2017;6(4):1 [FREE Full text] [doi: 10.1080/17513057.2013.769615]
78. Chinese online hospital embraces new opportunities: the market of online medical platform will exceed 50 billion RMB and people aged between 18 to 24 will be the main consumers. China Industrial Information. 2020 Jun 15. URL: https://www.chyxx.com/industry/202006/873996.html [accessed 2021-01-30]
79. Ayres J, Colby‐Rotell N, Wadleigh PM, Hopf T. Measuring patients' communication apprehension about interacting with physicians: instrument development. Commun Res Rep 1996 Mar;13(1):86-93. [doi: 10.1080/08824099609362074]
80. Wright KB, Frey L, Sopory P. Willingness to communicate about health as an underlying trait of patient self-advocacy: the development of the willingness to communicate about health (WTCH) measure. Commun Stud 2007 Feb 16;58(1):35-51. [doi: 10.1080/10510970601168673]
81. McCroskey JC. Reliability and validity of the willingness to communicate scale. Commun Q 1992 Jan;40(1):16-25. [doi: 10.1080/01463379209369817]
82. 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]
83. Hayes AF. PROCESS: a versatile computational tool for observed variable mediation, moderation, and conditional process modeling. 2012. URL: http://www.afhayes.com/public/process2012.pdf [accessed 2021-04-02]
84. Donovan LA, MacIntyre PD. Age and sex differences in willingness to communicate, communication apprehension, and self-perceived competence. Commun Res Rep 2004 Sep;21(4):420-427. [doi: 10.1080/08824090409360006]
85. Teven JJ, Richmond VP, McCroskey JC, McCroskey LL. Updating relationships between communication traits and communication competence. Commun Res Rep 2010 Aug 04;27(3):263-270. [doi: 10.1080/08824096.2010.496331]
86. Croucher SM. Communication apprehension, self-perceived communication competence, and willingness to communicate: a French analysis. J Int Intercult Commun 2013 Nov;6(4):298-316. [doi: 10.1080/17513057.2013.769615]
87. Baker SC, Watson BM. How patients perceive their doctors' communication. J Lang Soc Psychol 2015 May 18;34(6):621-639. [doi: 10.1177/0261927x15587015]
88. Allen R, Brown KL. Developing Communication Competence in Children: A Report of the Speech Communication Association's National Project on Speech Communication Competencies. Falls Church: Speech Communication Association; 1976.
89. Rice L, Markey PM. The role of extraversion and neuroticism in influencing anxiety following computer-mediated interactions. Pers Indiv Diff 2009 Jan;46(1):35-39. [doi: 10.1016/j.paid.2008.08.022]

Abbreviations

AVE: average variance extracted
CA: communication apprehension
CMC: computer-mediated communication
CR: construct reliability
PSP: perceived social presence
RMB: renminbi
SPCC: self-perceived communication competence
WTC: willingness to communicate
WTCH: willingness to communicate about health

Edited by A Mavragani; submitted 06.07.21; peer-reviewed by X Lu, S Atanasova; comments to author 18.08.21; revised version received 30.10.21; accepted 14.04.22; published 11.05.22.

Please cite as:
Chen L, Zhang D, Hou M
The Influence of Perceived Social Presence on the Willingness to Communicate in Mobile Medical Consultations: Experimental Study
J Med Internet Res 2022;24(5):e31797
URL: https://www.jmir.org/2022/5/e31797
doi:10.2196/31797
PMID:35544293

XSL•FO
RenderX

©Lijuan Chen, Danyang Zhang, Mutian Hou. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.05.2022. 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.

XSL•FO
RenderX

Original Paper

Screening Depressive Symptoms and Incident Major Depressive Disorder Among Chinese Community Residents Using a Mobile App–Based Integrated Mental Health Care Model: Cohort Study

Huimin Zhang[1,2*], PhD; Yuhua Liao[1,2*], PhD; Xue Han[2], PhD; Beifang Fan[2], MD; Yifeng Liu[2], MA; Leanna M W Lui[3,4,5], PhD; Yena Lee[3,4,5], PhD; Mehala Subramaniapillai[3,4,5], PhD; Lingjiang Li[6], MD, PhD; Lan Guo[1*], MD, PhD; Ciyong Lu[1*], MD, PhD; Roger S McIntyre[3,4,5], MD, PhD

[1]Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University, Guangzhou, China

[2]Department of Psychiatry, Shenzhen Nanshan Center for Chronic Disease Control, Shenzhen, China

[3]Mood Disorders Psychopharmacology Unit, University Health Network, Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[4]Institute of Medical Science, University of Toronto, Toronto, ON, Canada

[5]Department of Pharmacology, University of Toronto, Toronto, ON, Canada

[6]Mental Health Institute of the Second Xiangya Hospital, Central South University, Changsha, China

[*]these authors contributed equally

Corresponding Author:
Ciyong Lu, MD, PhD
Department of Medical Statistics and Epidemiology
School of Public Health
Sun Yat-sen University
No.74, Zhongshan 2nd
Guangzhou, 510080
China
Phone: 86 020 87332477
Email: luciyong@mail.sysu.edu.cn

Abstract

Background: Depression is associated with significant morbidity and human capital costs globally. Early screening for depressive symptoms and timely depressive disorder case identification and intervention may improve health outcomes and cost-effectiveness among affected individuals. China's public and academic communities have reached a consensus on the need to improve access to early screening, diagnosis, and treatment of depression.

Objective: This study aims to estimate the screening prevalence and associated factors of subthreshold depressive symptoms among Chinese residents enrolled in the cohort study using a mobile app–based integrated mental health care model and investigate the 12-month incidence rate and related factors of major depressive disorder (MDD) among those with subthreshold depressive symptoms.

Methods: Data were drawn from the Depression Cohort in China (DCC) study. A total of 4243 community residents aged 18 to 64 years living in Nanshan district, Shenzhen city, in Guangdong province, China, were encouraged to participate in the DCC study when visiting the participating primary health care centers, and 4066 (95.83%) residents who met the DCC study criteria were screened for subthreshold depressive symptoms using the Patient Health Questionnaire-9 at baseline. Of the 4066 screened residents, 3168 (77.91%) with subthreshold depressive symptoms were referred to hospitals to receive a psychiatric diagnosis of MDD within 12 months. Sleep duration, anxiety symptoms, well-being, insomnia symptoms, and resilience were also investigated. The diagnosis of MDD was provided by trained psychiatrists using the Mini-International Neuropsychiatric Interview. Univariate and multivariate logistic regression models were performed to explore the potential factors related to subthreshold depressive symptoms at baseline, and Cox proportional hazards models were performed to explore the potential factors related to incident MDD.

Results: Anxiety symptoms (adjusted odds ratio [AOR] 1.63, 95% CI 1.42-1.87) and insomnia symptoms (AOR 1.13, 95% CI 1.05-1.22) were associated with an increased risk of subthreshold depressive symptoms, whereas well-being (AOR 0.93, 95% CI 0.87-0.99) was negatively associated with depressive symptoms. During the follow-up period, the 12-month incidence rate of

XSL•FO
RenderX

MDD among participants with subthreshold depressive symptoms was 5.97% (189/3168). After incorporating all significant variables from the univariate analyses, the multivariate Cox proportional hazards model reported that a history of comorbidities (adjusted hazard ratio [AHR] 1.49, 95% CI 1.04-2.14) and anxiety symptoms (AHR 1.13, 95% CI 1.09-1.17) were independently associated with an increased risk of incident MDD. The 5-item World Health Organization Well-Being Index was associated with a decreased risk of incident MDD (AHR 0.90, 95% CI 0.86-0.94).

Conclusions: Elevated anxiety symptoms and unfavorable general well-being were significantly associated with subthreshold depressive symptoms and incident MDD among Chinese residents in Shenzhen. Early screening for subthreshold depressive symptoms and related factors may be helpful for identifying populations at high risk of incident MDD.

(J Med Internet Res 2022;24(5):e30907) doi:10.2196/30907

KEYWORDS

screening; depressive symptoms; incident major depressive disorder; Chinese community residents; electronic-based integrated mental health care model

Introduction

Background

Mental disorders account for significant illness-associated disability globally, and major depressive disorder (MDD, also called clinical depression) is one of the leading causes [1]. Moreover, MDD is highly prevalent, with high recurrence rates, nonrecovery, chronicity, and interepisodic dysfunction [2]. Along with staggering human costs, MDD exacts enormous individual and societal costs [3]. It is reported that MDD is closely associated with a loss of productivity and noticeable personal, social, and economic decline, thereby creating significant demands on patients, families, society, and service providers [4]. Moreover, at its worst, MDD can lead to increased risks of suicidal behavior (eg, suicidal ideation, suicide attempts, or even suicide death), and a large proportion of depressed individuals have suicidal ideation and suicide attempts [5-7]. According to the World Health Organization (WHO), depression affects >300 million people and represents a major contributor to the global burden of disease [8].

It is well documented that subthreshold depressive symptoms (ie, not meeting the minimum diagnostic threshold for a major depressive episode) could predispose and portend incident MDD, and previous evidence suggests that individuals manifesting subthreshold depressive symptoms have an approximately 2-fold higher risk of incident MDD than those without [9,10]. The increasing prevalence of MDD and its associated impact on human function is a national health priority for all countries, notably societies and health care systems in most low- and middle-income countries (LMIC) [8]. The additional challenge in LMIC is the observation that an estimated <10% of individuals with MDD in LMIC receive minimal treatment and support services [11]. Moreover, the gap in implementation of clinical practice guidelines for MDD is greater in LMIC than in high-income countries [12]. Early screening for subthreshold depressive symptoms has been reported to increase the likelihood of case identification among affected individuals, and a positive screen for subthreshold depressive symptoms is suggested to trigger an additional diagnostic assessment and, thus, improve future health [13]. The American Academy of Family Physicians and US Preventive Services Task Force recommend screening for depression in general adults [14]. Moreover, previous evidence indicated that early diagnosis and treatment of clinical depressive disorder might result in better outcomes [12] and seem to be more cost-effective [15]. In addition, early screening for depressive symptoms and depressive disorders has the potential to be effective. However, this had not been previously established.

During the past 30 years across China, rapid economic development and social change (eg, urbanization) have exposed citizens to changing factors. These rapid changes may be a determinant of adverse mental health problems (eg, subthreshold depressive symptoms and MDD) [16]. In keeping with this view, an increasing rate of depressive symptoms and mood disorders has been reported in China. For example, a recent national study using data from the China Mental Health Survey among Chinese adults reported that the weighted 12-month prevalence and lifetime prevalence of depressive disorder were 3.6% and 6.8%, respectively. Using the 2012 China Family Panel Studies data, a separate survey reported that 37.86% of the adult respondents experienced depressive symptoms [16]. Previous analyses from our group using data from the School-based Chinese Adolescents Health Survey reported a high prevalence (5.6% to 8.3%) of depressive symptoms among Chinese adolescents [17-19]. Moreover, a previous meta-analysis reported that the pooled lifetime prevalence of suicidal ideation and that of suicide attempts among patients with MDD in China were 53.1% and 23.7%, respectively [7]. However, access to mental health care in China remains constrained and needs to be improved [20].

Cohort Study

Although 6 types of serious mental health disorders (including schizophrenia, schizoaffective disorder, persistent delusional disorder, bipolar disorder, mental disorders caused by epilepsy, and mental retardation accompanied by mental disorders) are recognized in community-based mental health management programs in China, MDD is not included. Previous evidence suggests that factors such as stigma-induced stress contribute to the unwillingness to seek professional help among individuals with depressive symptoms or MDD [21]. The increasing demand for mental health services and the shortage of psychiatrists in China have received the attention of Chinese policy makers and health care professionals [22]. Accordingly, China's public and academic communities have reached a consensus on the need to improve access to early screening, diagnosis, and treatment

XSL•FO
RenderX

of depression [23]. However, few studies have been conducted in China related to the screening and prevention of subthreshold depressive symptoms and MDD in community residents and the development of integrated mental health care models connecting primary, hospital, and community care divisions. Therefore, we performed this cohort study to estimate the screening prevalence and related factors of subthreshold depressive symptoms among community residents in Shenzhen, Guangdong province, China, using a self-developed mobile app–based integrated mental health care model and determine the 12-month incidence rate and related factors of incident MDD among individuals with depressive symptoms.

Methods

Study Design

Data were derived from the Depression Cohort in China (DCC) study (Chinese Clinical Trial Registry ChiCTR 1900022145), which is an ongoing longitudinal, population-based study for early identification, treatment, prevention, and management of subthreshold depressive symptoms and MDD [24]. We developed a Toronto-based Building Bridges to Integrate Care (BRIDGES) health care model to standardize the screening, detection, and treatment of individuals with subthreshold depressive symptoms or MDD in Nanshan district, Shenzhen, to meet the mental health needs of residents and their families [25]. With a population of approximately 2 million in an area of 185 km^2, Nanshan is one of the most densely populated districts of Shenzhen.

Ethics Approval

The study procedures were carried out in accordance with the Declaration of Helsinki. This study received ethics approval from the institutional review board of the School of Public Health, Sun Yat-sen University (L2017044), and the study protocol was approved by the ethics review boards of all the participating centers.

BRIDGES Health Care Model

The DCC study used a BRIDGES health care model, which used the BRIDGES model, a project of the University of Toronto's departments of medicine and family and community medicine, as a reference [25], to link mental health care delivery among primary health care centers, a general hospital, and a specialized mental health hospital in accordance with the health system in Nanshan. In the integrated health care model, individuals are screened at primary health care centers by general practitioners (GPs) at baseline and there is a referral gateway between primary health care centers and general and specialized mental health hospitals in the DCC study. Those who screen positive with subthreshold depressive symptoms at primary health care centers will be referred to general or specialized mental health hospitals to receive psychiatric diagnoses within a follow-up period of a maximum of 12 months. Participants referred through this gateway do not need to go through the hospital patient registration process and are given priority for care at the hospital. Considering that almost all GPs are not professional psychiatrists in China, psychiatrists from specialist hospitals trained GPs at primary health care centers to identify subthreshold depressive symptoms and provide usual care, referral, and follow-up for participants with these symptoms. Psychiatrists at hospitals provided outpatient or hospitalized care and education programs to patients diagnosed with MDD as well as follow-up, management, and referrals. Moreover, in the DCC study, project managers, who were public health physicians from general hospitals, supervised and ensured the quality of our integrated health care implementation process. Notably, in the DCC study, a mobile phone app, which included screening, referral, follow-up, and management functions, was developed and used by the GPs from the primary health care centers, psychiatrists from participating specialist hospitals, and project managers. Besides, the eligible participants at the primary care centers would be provided an account number by the GPs to access the app to complete the screening questionnaire and follow-up assessment when they visit the primary health care centers and hospitals in the corresponding study stages. The study process in the DCC study was performed through the self-developed mobile app, and it has been reported that this digital data collection mechanism may be a promising tool to collect data related to other diseases and risk factors [26].

Participants

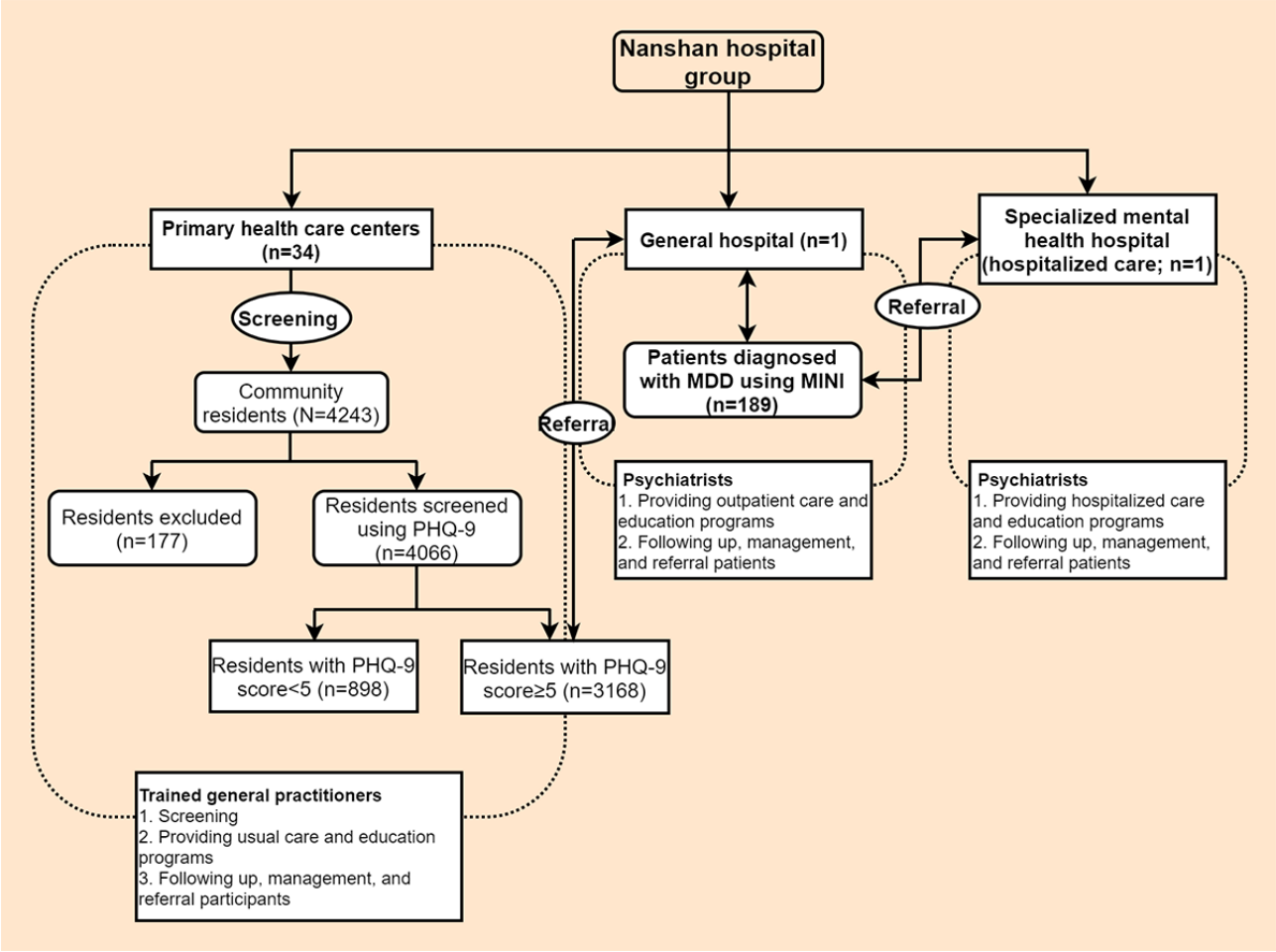
The study data were drawn from an ongoing cohort study that began in early 2019 in which community residents are screened when they visit primary health care centers. Approximately 90,000 residents in Nanshan district walk through the doors of 34 primary health care centers a year. Among all the people visiting these centers, GPs selectively screen those who have mental health–related physical complaints (eg, sleep problems and chronic somatic pain) or are more likely to have mental health issues based on the GPs' clinical experience and our study training. Our study aimed to screen individuals with subthreshold depressive symptoms and identify patients with MDD within limited medical resources and periods. Therefore, a total of 4243 community residents aged 18 to 64 years living in Nanshan were encouraged to participate in the DCC study when visiting the 34 participating primary health care centers at baseline, of which 177 (4.17%) residents were excluded (n=5, 2.8%, with incomplete information on depressive symptoms; n=133, 75.1%, with diagnostic depressive disorder; and n=39, 22%, with other psychiatric disorders; Figure 1), leaving 4066 (95.83%) residents who met the DCC study criteria and were screened for subthreshold depressive symptoms at baseline by the trained GPs at the participating primary health care centers. The DCC study exclusion criteria were as follows: (1) a diagnosis of current, or history of, depressive disorder, severe psychiatric disorder (ie, bipolar disorder, schizophrenia, schizoaffective mental disorder, paranoid mental disorder, mental disorders caused by epilepsy, or mental retardation), or alcohol or drug dependence disorder; (2) pregnant or perinatal women; (3) nonfluency in Mandarin; (4) inability to understand questionnaires or provide consent for themselves; (5) living outside the community; and (6) having a plan to leave Shenzhen within 12 months.

In this study, of the 4066 residents who met the study criteria, 3168 (77.91%) screened positive with subthreshold depressive symptoms at baseline at the primary health care centers and

XSL•FO
RenderX

were referred to the general or specialized mental health hospitals to receive the psychiatric diagnoses within 12 months through the BRIDGES health care model. Psychiatric diagnoses were provided by the trained psychiatrists using the Mini-International Neuropsychiatric Interview (MINI; Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria). Among the patients with subthreshold depressive symptoms, 5.97% (189/3168) were first diagnosed with MDD during the follow-up period after the baseline screening (Figure 1). Written informed consent explaining the study purposes, processes, benefits, and risks was obtained from each participant.

Figure 1. The integrated mental health care model in Nanshan, Shenzhen, in Guangdong province, China. MDD: major depressive disorder; MINI: Mini-International Neuropsychiatric Interview; PHQ-9: Patient Health Questionnaire-9.



Measures

Subthreshold Depressive Symptoms

In the DCC study, subthreshold depressive symptoms were measured using the Patient Health Questionnaire-9 (PHQ-9), a widely used self-report measure in clinical and research settings that screens for depressive symptoms over the past 2 weeks [27]. The Cronbach α for PHQ-9 was .80 in this study. The PHQ-9 consists of 9 items, each addressing specific symptoms of depression during the past 2 weeks, and the scores for each item range from *0=not at all* to *3=nearly every day*, with a maximum score of 27. Higher scores were indicative of more severe depressive symptomatology. In this study, participants with a PHQ-9 score of ≥5 and without current, or a history of, depressive disorders were operationalized as having subthreshold depressive symptoms [24,28,29].

Ascertainment of Incident MDD

Participants with subthreshold depressive symptoms were referred to hospitals to receive the diagnosis of MDD within 12 months. The MINI, a Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision–based validated structured diagnostic psychiatric interview, was used by psychiatrists to diagnose a current MDD and exclude other diagnoses [30].

Independent Variables

Sleep duration was assessed by the question, *How many hours do you usually sleep each day?* Anxiety symptoms were assessed by the Generalized Anxiety Disorder Scale-7 [31], which has been validated and extensively used in Chinese studies with satisfactory psychometric properties [32]. The Cronbach α was .92 with our sample. The 7 items' total score ranges from 0 to 21, with higher scores indicating more severe anxiety symptomatology.

Well-being was measured using the 5-item World Health Organization Well-Being Index (WHO-5), which is a positively worded scale designed to measure the level of subjective well-being over the past 2 weeks on a 6-point scale ranging

XSL•FO
RenderX

from 0 (not present) to 5 (constantly present), leading to a raw score ranging from 0 (absence of well-being) to 25 (maximal well-being) [33]. The Cronbach α for WHO-5 was .94 in this study.

Insomnia symptoms were assessed with the Insomnia Severity Index, which consists of 7 items, with each item scored from 0 to 4, for a maximum of 28 points. Higher scores represent greater insomnia levels [34], and the Cronbach α was .93 with our sample.

Adverse life events were measured using the Stressful Life Events Screening Questionnaire (SLESQ), which has been validated in Chinese studies [35], and the Cronbach α was .74 in this study. The SLESQ includes 12 items, each with a dichotomous response option (0=no and 1=yes). Higher scores reflect the experience of more adverse life events.

Resilience was measured using the Connor-Davidson Resilience Scale (CD-RISC), which comprises 25 items, with each rated on a 5-point scale ranging from 0 (not at all true) to 4 (true nearly all of the time). The Cronbach α for CD-RISC was .95 in this study. The CD-RISC yields a total resilience score ranging from 0 to 100, with higher scores indicating greater resilience [36].

The sociodemographic variables included in this study were age, sex (1=male and 2=female), ethnicity (1=Han Chinese and 2=Chinese minorities), education level (1=junior high school or below, 2=senior high school, and 3=college or above), living arrangement (1=living alone, 2=living with family, and 3=living with others), marital status (1=unmarried, 2=married, 3=divorced, and 4=widowed), lifetime smoking (assessed by the question, *Have you ever smoked a cigarette?* Responses were coded as 1=yes and 2=no) [37], onset age of smoking, lifetime drinking (assessed by the question, *Have you ever consumed at least one alcoholic drink of any kind?* Responses were coded as 1=yes and 2=no) [37-39], and onset age of drinking. History of comorbidities (including hypertension, diabetes, heart disease, stroke, thyroid disease, tumor, and others) were also collected (responses coded as 1=yes and 2=no).

Statistical Analysis

Data were described as means (SDs) for normally distributed continuous variables and as medians (IQRs) for nonnormally distributed continuous variables, and frequency with percentage was used to describe categorical variables. Baseline characteristics were summarized according to baseline depressive symptoms. Mann-Whitney *U* tests or 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables were conducted to compare baseline sample characteristics between participants with a PHQ-9 score of <5 and those with a PHQ score of ≥5, as appropriate. Univariate and multivariate logistic regression models were performed to explore the potential factors related to subthreshold depressive symptoms at baseline, and odds ratios (ORs) with 95% CIs were estimated. All the variables shown to be significantly associated with subthreshold depressive symptoms by the univariate logistic regression models were entered into the multivariate logistic regression models. Moreover, the 12-month incidence rate of MDD among participants with subthreshold depressive symptoms was calculated. Univariate and multivariate Cox proportional hazards models were performed to explore the potential factors related to incident MDD, and hazard ratios (HRs) with 95% CIs were also reported. All the variables shown to be significantly associated with incident MDD by the univariate Cox proportional hazards models were incorporated into the multivariate Cox proportional hazards models. In addition, we also explored the associations of observed risk factors with incident MDD using 3-knotted restricted cubic spline regression models, and the *P* values for the test of linearity hypotheses were reported. Moreover, regarding the *P* values calculated from the multivariate logistic regression models or multivariate Cox proportional hazards models, the false discovery rate was calculated to address the concern of potential type I errors and multiple hypotheses testing. The false discovery rate–adjusted *P* value was indicated by *q*, and the results were considered nominally significant when $q<.10$ [40]. The multiple imputation by chained equations method was applied for missing data [41]. All statistical analyses were conducted using Stata software (version 14.1; StataCorp LLC) and R statistical software (version 4.0.2; R Foundation for Statistical Computing). All statistical tests were 2-sided, and $P<.05$ was considered statistically significant.

Results

Baseline Sample Characteristics

The sample characteristics of all included participants at baseline are shown in Table 1. Among the 4066 participants, the mean age was 38.19 (SD 11.46) years, and 1541 (37.9%) were men; 576 (14.17%) reported an education level as junior high school or below; 484 (11.9%) reported living alone; 1080 (26.56%) reported lifetime smoking; 2424 (59.61%) reported lifetime drinking; and 974 (23.95%) reported having a history of comorbidities. The mean (SD) values of sleep duration, Generalized Anxiety Disorder Scale-7 score, WHO-5 score, Insomnia Severity Index score, SLESQ score, and CD-RISC score were 6.67 (4.68), 5.24 (4.88), 13.75 (6.05), 8.25 (6.54), 0.48 (1.15), and 59.05 (23.50), respectively. Of the 4066 participants, 3168 (77.91%) had a PHQ-9 score of ≥5. The differences between the groups with and without subthreshold depressive symptoms were not significant regarding the distribution of age, sex, ethnicity, lifetime smoking, and onset age of smoking. The characteristics of each item of the PHQ-9 among participants at baseline are presented in Multimedia Appendix 1.

XSL•FO
RenderX

Table 1. Baseline characteristics of participants according to subthreshold depressive symptom status (N=4066).

Variable	Participants			*P* value[a]
	Total	PHQ-9 score<5 (n=898)	PHQ-9 score≥5 (n=3168)	
Age (years), mean (SD)	38.19 (11.46)	38.35 (10.74)	38.14 (11.66)	.64
Sex, n (%)				.67
Male	1541 (37.9)	346 (38.53)	1195 (37.72)	
Female	2525 (62.1)	552 (61.47)	1973 (62.28)	
Ethnicity, n (%)				.76
Han Chinese	3922 (96.46)	869 (96.77)	3053 (96.37)	
Chinese minorities	140 (3.44)	29 (3.23)	111 (3.50)	
Missing	4 (0.1)	—[b]	—	
Education level, n (%)				<.001
Junior high school or below	576 (14.17)	84 (9.35)	492 (15.53)	
Senior high school	966 (23.76)	198 (22.05)	768 (24.24)	
College or above	2516 (61.88)	615 (68.49)	1901 (60)	
Missing	8 (0.2)	—	—	
Living arrangement, n (%)				<.001
Living alone	484 (11.9)	83 (9.24)	401 (12.66)	
Living with family	3018 (74.23)	723 (80.51)	2295 (72.44)	
Living with others	447 (11)	69 (7.68)	378 (11.93)	
Missing	117 (2.88)	—	—	
Marital status, n (%)				<.001
Unmarried	997 (24.52)	178 (19.82)	819 (25.85)	
Married	2927 (72)	704 (78.4)	2223 (70.17)	
Divorced	117 (2.88)	13 (1.45)	104 (3.28)	
Widowed	25 (0.61)	3 (0.33)	22 (0.69)	
Lifetime smoking (yes), n (%)	1080 (26.56)	225 (25.06)	855 (27)	.27
Onset age of smoking (years), mean (SD)	19.50 (5.39)	19.52 (6.20)	19.49 (5.15)	.93
Lifetime drinking (yes), n (%)	2424 (59.61)	566 (63.02)	1858 (58.65)	.02
Onset age of drinking (years), mean (SD)	19.23 (4.93)	19.72 (4.90)	19.09 (4.94)	.008
History of comorbidities (yes), n (%)	974 (23.95)	189 (21.05)	785 (24.78)	.02
Hypertension	396 (9.74)	84 (9.35)	312 (9.85)	.70
Diabetes	168 (4.13)	37 (4.12)	131 (4.14)	.99
Heart disease	45 (1.11)	8 (0.89)	37 (1.17)	.59
Stroke	13 (0.32)	1 (0.11)	12 (0.38)	.32
Thyroid disease	129 (3.17)	26 (2.9)	103 (3.25)	.67
Tumor	33 (0.81)	4 (0.45)	29 (0.91)	.21
Other	313 (7.7)	52 (5.79)	261 (8.24)	.02
Sleep duration (hours per day), mean (SD)	6.67 (4.68)	7.25 (6.54)	6.50 (3.99)	<.001
Anxiety symptoms, mean (SD)	5.24 (4.88)	1.37 (1.86)	6.34 (4.91)	<.001
Adverse life events, mean (SD)	0.48 (1.15)	0.18 (0.65)	0.56 (1.24)	<.001
Well-being, mean (SD)	13.75 (6.05)	12.45 (5.80)	18.35 (4.49)	<.001
Insomnia symptoms, mean (SD)	8.25 (6.54)	3.45 (3.47)	9.61 (6.56)	<.001

XSL•FO
RenderX

Variable	Participants			*P* value[a]
	Total	PHQ-9 score<5 (n=898)	PHQ-9 score≥5 (n=3168)	
Resilience, mean (SD)	59.05 (23.50)	68.56 (26.36)	56.05 (21.69)	<.001

[a]Mann-Whitney *U* test or 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables were conducted to compare baseline sample characteristics between participants with and without subthreshold depressive symptoms, as appropriate.

[b]Not available.

Factors Associated With Subthreshold Depressive Symptoms

Univariate logistic regression models reported that participants with education levels of junior high school or below (OR 1.26, 95% CI 1.05-1.50) and senior high school (OR 1.90, 95% CI 1.48-2.43) had higher risks of having subthreshold depressive symptoms than those with education level of college or above. Participants living with family (OR 0.66, 95% CI 0.51-0.84) were less likely to report subthreshold depressive symptoms than those living alone. Lifetime drinking (OR 1.20, 95% CI 1.03-1.40) and a history of comorbidities (OR 1.24, 95% CI 1.03-1.48) were positively associated with subthreshold depressive symptoms, as were anxiety symptoms (OR 1.78, 95% CI 1.70-1.87), insomnia symptoms (OR 1.30, 95% CI 1.27-1.32), and adverse life events (OR 1.73, 95% CI 1.52-1.96). The onset age of drinking (OR 0.98, 95% CI 0.96-0.99), sleep duration (OR 0.88, 95% CI 0.83-0.94), general well-being (OR 0.81, 95% CI 0.79-0.82), and resilience (OR 0.98, 95% CI 0.97-0.98) were negatively associated with subthreshold depressive symptoms.

After incorporating all significant variables from the univariate analyses, the multivariate logistic regression model demonstrated that only anxiety symptoms (adjusted OR [AOR] 1.63, 95% CI 1.42-1.87) and insomnia symptoms (AOR 1.13, 95% CI 1.05-1.22) were associated with an increased risk of subthreshold depressive symptoms. General well-being (AOR 0.93, 95% CI 0.87-0.99) was negatively associated with the risk of subthreshold depressive symptoms. Moreover, these factors were still significantly associated with subthreshold depressive symptoms after correcting for multiple testing (Table 2).

XSL•FO
RenderX

Table 2. Factors associated with subthreshold depressive symptoms among baseline participants.

Variable	Model 1[a]		Model 2[b]		
	OR[c] (95% CI)	*P* value	Adjusted OR (95% CI)	*P* value	*q*[d]
Age (1-year increase)	1.00 (0.99-1.01)	.64	N/A[e]	N/A	N/A
Male (reference=female)	0.97 (0.83-1.13)	.66	N/A	N/A	N/A
Ethnicity (reference=Chinese minorities)	0.92 (0.61-1.39)	.69	N/A	N/A	N/A
Education level (reference=college or above)					
Junior high school or below	1.26 (1.05-1.50)	.01	1.48 (0.57-3.88)	.43	.78
Senior high school	1.90 (1.48-2.43)	<.001	1.13 (0.54-2.38)	.75	.89
Living arrangement (reference=living alone)					
Living with family	0.66 (0.51-0.84)	.001	0.92 (0.30-2.84)	.89	.96
Living with others	1.13 (0.80-1.61)	.48	0.54 (0.24-1.21)	.13	.39
Marital status (reference=widowed)					
Unmarried	0.63 (0.19-2.12)	.45	N/A	N/A	N/A
Married	0.43 (0.13-1.44)	.17	N/A	N/A	N/A
Divorced	1.09 (0.29-4.15)	.90	N/A	N/A	N/A
Lifetime smoking (reference=no smoking)	1.11 (0.93-1.31)	.25	N/A	N/A	N/A
Onset age of smoking (1-year increase)	1.00 (0.97-1.03)	.93	N/A	N/A	N/A
Lifetime drinking (reference=no drinking)	1.20 (1.03-1.40)	.02	1.33 (0.39-4.59)	.65	.89
Onset age of drinking (1-year increase)	0.98 (0.96-0.99)	.009	0.96 (0.91-1.01)	.15	.39
History of comorbidities (reference=no comorbidities)	1.24 (1.03-1.48)	.02	1.26 (0.67-2.36)	.48	.78
Sleep duration (1-hour increase)	0.88 (0.83-0.94)	<.001	0.98 (0.94-1.02)	.33	.71
Anxiety symptoms (increase in score by 1)	1.78 (1.70-1.87)	<.001	1.63 (1.42-1.87)	<.001	<.001
Well-being (increase in score by 1)	0.81 (0.79-0.82)	<.001	0.93 (0.87-0.99)	.02	.09
Insomnia symptoms (increase in score by 1)	1.30 (1.27-1.32)	<.001	1.13 (1.05-1.22)	.001	.007
Adverse life events (increase in score by 1)	1.73 (1.52-1.96)	<.001	0.96 (0.74-1.24)	.76	.89
Resilience (increase in score by 1)	0.98 (0.97-0.98)	<.001	1.00 (0.99-1.01)	.99	.99

[a]The univariate logistic regression models were the unadjusted models.

[b]The multivariate logistic regression models incorporated all significant variables from the univariate analyses.

[c]OR: odds ratio.

[d]The false discovery rate–adjusted *P* value.

[e]N/A: not applicable.

Factors Associated With Incident MDD Among Participants With Subthreshold Depressive Symptoms

Of the 3168 residents screened with subthreshold depressive symptoms at baseline, 189 (5.97%) met the first major depressive episode criterion between March 2019 and March 2020; the 12-month incidence rate of MDD among participants with subthreshold depressive symptoms was 5.97% (189/3168; Multimedia Appendix 2). Table 3 highlights the factors associated with incident MDD. The univariate Cox proportional hazards models reported that lifetime drinking (HR 1.51, 95% CI 1.10-2.06), a history of comorbidities (HR 2.05, 95% CI 1.44-2.91), anxiety symptoms (HR 1.24, 95% CI 1.21-1.27), insomnia symptoms (HR 1.15, 95% CI 1.13-1.18), and adverse life events (HR 1.37, 95% CI 1.28-1.47) were positively associated with elevated risks of incident MDD. General well-being (HR 0.80, 95% CI 0.78-0.83) and resilience (HR 0.98, 95% CI 0.97-0.99) were negatively associated with incident MDD. After incorporating all significant variables from the univariate analyses, the multivariate Cox proportional hazards models demonstrated that a history of comorbidities was independently associated with a 49% increased risk of incident MDD (adjusted HR [AHR] 1.49, 95% CI 1.04-2.14) and anxiety symptoms (AHR 1.13, 95% CI 1.09-1.17) were positively associated with incident MDD. General well-being was associated with a decreased risk of incident MDD (AHR 0.90, 95% CI 0.86-0.94). Moreover, these associations were still significant after correcting for multiple testing.

In addition, we used restricted cubic splines to flexibly model and visualize the associations of anxiety symptoms and well-being with the risk of incident MDD (Multimedia Appendix

XSL•FO
RenderX

3). A linear and positive association between the anxiety symptoms' total score and risk of incident MDD was also found (*P* for nonlinearity=.90), and a nonlinear and negative association between the well-being scores and risk of incident MDD was observed (*P* for nonlinearity=.01).

Table 3. Factors associated with incident major depressive disorder among participants with subthreshold depressive symptoms.

Variable	Model 1[a]		Model 2[b]		
	HR[c] (95% CI)	*P* value	Adjusted HR (95% CI)	*P* value	*q*[d]
Age (1-year increase)	1.00 (0.99-1.02)	.73	N/A[e]	N/A	N/A
Male (reference=female)	0.77 (0.56-1.05)	.10	N/A	N/A	N/A
Ethnicity (reference=Chinese minorities)	0.73 (0.36-1.50)	.40	N/A	N/A	N/A
Education level (reference=college or above)					
Junior high school or below	0.86 (0.55-1.33)	.50	N/A	N/A	N/A
Senior high school	1.00 (0.70-1.42)	.99	N/A	N/A	N/A
Living arrangement (reference=living alone)					
Living with families	0.70 (0.46-1.06)	.01	N/A	N/A	N/A
Living with others	1.00 (0.59-1.71)	.99	N/A	N/A	N/A
Marital status (reference=widowed)					
Unmarried	0.48 (0.14-1.67)	.25	N/A	N/A	N/A
Married	0.33 (0.08-1.12)	.07	N/A	N/A	N/A
Divorced	0.73 (0.19-2.78)	.64	N/A	N/A	N/A
Lifetime smoking (reference=no smoking)	1.02 (0.73-1.40)	.93	N/A	N/A	N/A
Onset age of smoking (1-year increase)	1.00 (0.96-1.04)	.98	N/A	N/A	N/A
Lifetime drinking (reference=no drinking)	1.51 (1.10-2.06)	.01	0.98 (0.68-1.42)	.92	.92
Onset age of drinking (1-year increase)	0.99 (0.96-1.02)	0.39	N/A	N/A	N/A
History of comorbidities (reference=no comorbidities)	2.05 (1.44-2.91)	<.001	1.49 (1.04-2.14)	.03	.07
Sleep duration (1-hour increase)	1.01 (0.99-1.03)	.36	N/A	N/A	N/A
Anxiety symptoms (increase in score by 1)	1.24 (1.21-1.27)	<.001	1.13 (1.09-1.17)	<.001	<.001
Well-being (increase in score by 1)	0.80 (0.78-0.83)	<.001	0.90 (0.86-0.94)	<.001	<.001
Insomnia symptoms (increase in score by 1)	1.15 (1.13-1.18)	<.001	1.03 (1.00-1.07)	.05	.07
Adverse life events (increase in score by 1)	1.37 (1.28-1.47)	<.001	1.09 (1.00-1.19)	.05	.07
Resilience (increase in score by 1)	0.98 (0.97-0.99)	<.001	1.00 (0.99-1.02)	.67	.78

[a]The univariate logistic regression models were the unadjusted models.

[b]The multivariate logistic regression models incorporated all significant variables from the univariate analyses.

[c]HR: hazard ratio.

[d]The false discovery rate–adjusted *P* value.

[e]N/A: not applicable.

Discussion

Principal Findings

This prospective cohort study used a mobile app–based integrated mental health care model to link mental health care delivery among primary health care centers, a general hospital, and a mental health hospital in Nanshan, Shenzhen, and identify populations at high risk and factors contributing to elevated risks of subthreshold depressive symptoms and incident MDD among Chinese residents in Nanshan.

Of the 4066 community residents meeting the DCC study criteria, 3168 (77.91%) screened positive for subthreshold depressive symptoms at baseline in evaluations by GPs using the PHQ-9 at primary health care centers [42]. This rate was higher than the prevalence reported in a previous study among adults in mainland China aged ≥45 years between 2011 and 2012 (26%) [43] and in a study among community people with ≥1 chronic conditions in Hong Kong between 2009 and 2011 (17%) [29]. The aforementioned differences may be attributed to the use of different scales. Another explanation for these results may be that the rapid economic growth and social change in recent years were accompanied by a general increase in

XSL•FO
RenderX

psychological pressure and stress in Shenzhen, one of the fastest-growing cities in China [44]. In addition, the higher screening rate observed in this study might also be explained by the successfully implemented integrated mental health care model. It means that the participants in the DCC study were not randomly selected, and they were invited for subthreshold depressive symptoms screening when they visited the participating primary health care centers for some physical health problems (eg, somatic and sleep problems), which were prevalent comorbidities in depressive symptoms and depressive disorder [45,46]. Moreover, most of the 4066 participants meeting the DCC study criteria had higher education levels (n=2516, 61.88%), were women (n=2525, 62.1%), and had lifetime drinking (n=2424, 59.61%) or a history of comorbidities (n=974, 23.95%), and these features had been reported to be possibly associated with depression development [29,47,48].

The univariate logistic regression models demonstrated that a lower level of education, lifetime drinking, a history of comorbidities, anxiety symptoms, insomnia symptoms, and adverse life events were positively associated with subthreshold depressive symptoms. In contrast, residents living with family, having an older onset age of drinking, having longer sleep duration, and having higher resilience were less likely to experience subthreshold depressive symptoms [49-52]. Our findings are consistent with the available evidence. The findings from the univariate analyses will be helpful for identifying community residents who may be at risk of subthreshold depressive symptoms. We should focus on high-risk groups who present with the aforementioned adverse characteristics. Although some previous evidence suggested that higher education levels were positively associated with an increased risk of depressive symptoms [47], others reported that depression was significantly more prevalent among those with a low education level [53]. These mixed results may be related to the different classification of education levels, the variety in sample characteristics (eg, age or biological gender), or the different socioeconomic environments. In this study, the observed finding of the unadjusted association between education level and subthreshold depressive symptoms may be related to the possibility that individuals with a lower education level in Shenzhen were more likely to struggle in their lives than those having an education level of college or above; therefore, they might be more likely to contend with emotional disturbance. Moreover, after incorporating all significant variables into the multivariate logistic regression models, the results showed that only anxiety and insomnia symptoms were significantly associated with an increased risk of subthreshold depressive symptoms, whereas general well-being was negatively associated with a risk of subthreshold depressive symptoms in this community sample. These findings may indicate that anxiety or insomnia symptoms are the core factors that influence the risk of subthreshold depressive symptoms. These symptoms may be the most important modifiable risk factors, and specific attention should be paid to populations experiencing anxiety or insomnia symptoms. It has been reported that anxiety and depressive symptoms overlap in various domains. For example, negative emotions and cognitive distortions may be the core causes or symptoms of anxiety and depressive symptoms, with differences in terms of severity [54]. Long-term anxiety symptoms are likely to lead to the onset of depressive symptoms [55]. A non–mutually exclusive explanation for the association between insomnia symptoms and depressive symptoms is sleep loss, resulting in cognitive and emotional impairments through the hyperactivity of the hypothalamic-pituitary-adrenal axis or increasing levels of inflammatory markers, which are possible common pathophysiological mechanisms of subthreshold depressive symptoms [56,57].

Regarding the situation of MDD in China, Huang et al [44] reported that the weighted 12-month prevalence of MDD was 3.6% among Chinese households between 2013 and 2015; Chen et al [58] found that the incidence of MDD was 4% among Chinese university students between 2007 and 2008. Taken together, a novel finding of this cohort study is the observed higher 12-month incidence rate of incident MDD among Chinese residents with subthreshold depressive symptoms (189/3168, 5.97%). The increased rate of incident MDD reported in this study may be attributed to the increased risks of developing a depressive disorder among individuals with subthreshold depressive symptoms compared with the general population [9,10]. Another explanation might be related to the fact that the follow-up period of some participants in this study occurred during the COVID-19 pandemic. The emergence of this global event has created an environment where many determinants of poor mental health are exacerbated, and depressive disorders had increased globally in 2020 because of the COVID-19 pandemic [59]. Our previous study using data from the DCC study also reported that the COVID-19 pandemic had a highly significant and negative impact on a population with subthreshold depressive symptoms [24]. In addition, this observed incidence rate of MDD might also indicate that implementing an app-based integrated mental health care model might be helpful for early detection of populations at high risk of the first episode of MDD.

Moreover, the univariate Cox proportional hazards models showed that lifetime drinking, a history of comorbidities, anxiety symptoms, insomnia symptoms, and adverse life events might predict an increased risk of incident MDD. A higher level of subjective well-being and resilience may predict a decreased risk of incident MDD. Findings from the univariate analyses may provide evidence for identifying populations at high risk for incident MDD and modifiable factors among individuals with subthreshold depressive symptoms. In addition, after accounting for all significant variables, the multivariate analyses indicated that only a history of comorbidities and anxiety symptoms were associated with an increased risk of incident MDD among populations with subthreshold depressive symptoms; a higher level of well-being significantly predicted decreased incident MDD risk. Furthermore, restricted cubic spline models demonstrated a linear and positive association between anxiety symptoms and the risk of incident MDD. Well-being was negatively associated with incident MDD in a nonlinear fashion, meaning that although individuals with lower general well-being might be at a higher risk of incident MDD, whereas those with a higher level of well-being might be less likely to develop MDD, the HR for incident MDD did not linearly decrease by the level of well-being. These findings suggest that recognizing and preventing individuals with a

XSL•FO
RenderX

history of comorbidities or anxiety symptoms from developing MDD may be the focus of targeted intervention efforts, and a strategy of cultivating well-being might be a promising first step. A possible explanation for the association between a history of comorbidities and incident MDD is that depressive disorder is prevalent in patients with a physical disorder (particularly in those with severe conditions such as diabetes and stroke), and this comorbidity largely contributes to a poorer quality of life, worsening outcomes, higher medical costs, and more significant disability of the physical disorders [60]. Similar to the results of subthreshold depressive symptoms, a significant association between anxiety symptoms and incident MDD was examined. Comorbid anxiety symptoms are common in patients with depressive disorder, and it has been widely reported that these disorders may share common underlying pathophysiology [61]. Moreover, the observed protective effects of well-being on incident MDD may be explained by its effects on positive psychological functioning, capturing one's level of positive life satisfaction and a sense of purpose in life [62]. Besides, a model promoted by Keyes [63] also implied that individuals experiencing many psychopathology symptoms were more likely to experience a low level of well-being and vice versa. Previous longitudinal studies have also shown the predictive value of well-being, specifically on depressive disorders [64]. Moreover, a previous study also provided possible evidence that supporting an eHealth intervention using a mobile app designed to improve the well-being of adults may be helpful for treating depressive symptomatology [65]. Hitherto, our findings suggest that targeted interventions to increase well-being may be effective in protecting against the risks of developing a depressive disorder.

Limitations

Several limitations need to be addressed. First, only community residents in Nanshan, Shenzhen, were involved in this study; thus, the findings may not be fully generalizable to other regions. Second, the study sample was drawn from the DCC study, which recruited participants from the participating primary health care centers, and the study sample was not randomly selected. Therefore, this study had selection and sampling bias and the estimated screening prevalence of depressive symptoms among adults in Shenzhen might be overestimated. Third, we did not estimate the 12-month incidence rate of MDD among individuals without subthreshold depressive symptoms. Although it may be rare for individuals without depressive symptoms to exhibit a 12-month incidence of MDD, these populations may present different illness characteristics in the presence of MDD. Fourth, the variable of PHQ-9 was used as a dichotomous variable (ie, having or not having depressive symptoms) in this study, and we would like to use a different method to estimate the severity of depressive symptoms (ie, the polytomous variable of PHQ-9) in our future study. To reduce the risk of developing MDD, early screening of vulnerable populations and implementation of effective interventions targeting these symptoms are highly recommended. The strengths of this study included the longitudinal design, the large representative community-based sample, and the use of a clinically validated diagnostic interview (ie, MINI) to diagnose MDD.

Conclusions

Using a mobile app–based integrated mental health care model, this study found that the screened prevalence of subthreshold depressive symptoms among community residents in Nanshan, Shenzhen, was high. More specifically, we reported that 5.97% (189/3168) of the individuals with subthreshold depressive symptoms developed MDD within 12 months. In addition, anxiety symptoms were associated with an increased risk of subthreshold depressive symptoms and incident MDD among the community residents, and the presence of a history of comorbidities may predict the elevated risk of incident MDD. Moreover, a higher level of general well-being might decrease the risks of subthreshold depressive symptoms and incident MDD. The results from our study highlight the following: (1) the 12-month incidence rate of MDD among populations with subthreshold depressive symptoms is high, and screening earlier on in the illness trajectory of individuals with subthreshold depressive symptoms and recognizing high-risk factors may lead to earlier detection and treatment of MDD; (2) more attention should be paid to vulnerable populations with adverse characteristics (eg, anxiety symptoms, insomnia symptoms, or adverse life events); and (3) the implementation of an integrated mental health care model (ie, linking community, primary health care centers, and hospitals) in China might be helpful for training GPs to provide essential mental health services, improving community residents' access to mental health care as well as the timely referral and management of patients with MDD.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (2018YFC2000705) and the National Natural Science Foundation of China (81761128030 and 81903339). The authors wish to give particular thanks to all participating primary health care centers, hospitals, and community residents who made the study possible and gratefully acknowledge technical support from the School of Public Health, Sun Yat-sen University.

Authors' Contributions

LG and CL had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. All authors were responsible for the concept and design of the study as well as the acquisition, analysis, and interpretation of data. HZ, Y Liao, and XH drafted the manuscript. Critical revision of the manuscript for important intellectual content was carried out by LG, CL, Y Lee, Y Liu, LMWL, MS, LL, and RSM. HZ was responsible for the statistical analysis. LG and CL obtained funding. Administrative, technical, or material support was provided by LG, CL, BF, and RSM. The study was supervised by LG, CL, BF, and RSM.

XSL•FO
RenderX

Conflicts of Interest

RSM has received research grant support from Canadian Institutes of Health Research (CIHR)/Global Alliance for Chronic Diseases (GACD)/National Natural Science Foundation of China (NSFC); speaker and consultation fees from Lundbeck, Janssen, Alkermes, Neumora Therapeutics, Mitsubishi Tanabe, Purdue, Pfizer, Otsuka, Takeda, Neurocrine, Sunovion, Bausch Health, Axsome, Novo Nordisk, Kris, Sanofi, Eisai, Intra-Cellular, NewBridge Pharmaceuticals, Abbvie, and Atai Life Sciences. RSM is a CEO of Braxia Scientific Corp..

Multimedia Appendix 1
The characteristics of each item of the Patient Health Questionnaire-9 among participants at baseline.
[DOCX File , 17 KB - jmir_v24i5e30907_app1.docx]

Multimedia Appendix 2
Incidence of the first diagnosis of depressive disorder among participants with subthreshold depressive symptoms.
[DOCX File , 15 KB - jmir_v24i5e30907_app2.docx]

Multimedia Appendix 3
Restricted cubic spline models for the associations of (A) anxiety symptoms and (B) well-being with the risk of incident major depressive disorder.
[DOCX File , 118 KB - jmir_v24i5e30907_app3.docx]

References

1. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva, Switzerland: World Health Organization; Jan 01, 2017.
2. Burcusa SL, Iacono WG. Risk for recurrence in depression. Clin Psychol Rev 2007 Dec;27(8):959-985 [FREE Full text] [doi: 10.1016/j.cpr.2007.02.005] [Medline: 17448579]
3. McIntyre RS, Millson B, Power GS. Burden of treatment resistant depression (TRD) in patients with major depressive disorder in Ontario using Institute for Clinical Evaluative Sciences (ICES) databases: Economic burden and healthcare resource utilization. J Affect Disord 2020 Dec 01;277:30-38 [FREE Full text] [doi: 10.1016/j.jad.2020.07.045] [Medline: 32791390]
4. McLaughlin KA. The public health impact of major depression: a call for interdisciplinary prevention efforts. Prev Sci 2011 Dec;12(4):361-371 [FREE Full text] [doi: 10.1007/s11121-011-0231-8] [Medline: 21732121]
5. Sokero TP, Melartin TK, Rytsälä HJ, Leskelä US, Lestelä-Mielonen PS, Isometsä ET. Suicidal ideation and attempts among psychiatric patients with major depressive disorder. J Clin Psychiatry 2003 Sep;64(9):1094-1100. [doi: 10.4088/jcp.v64n0916] [Medline: 14628986]
6. Gramaglia C, Feggi A, Bergamasco P, Bert F, Gattoni E, Marangon D, et al. Clinical characteristics associated with suicide attempts in clinical settings: a comparison of suicidal and non-suicidal depressed inpatients. Front Psychiatry 2016;7:109 [FREE Full text] [doi: 10.3389/fpsyt.2016.00109] [Medline: 27378957]
7. Dong M, Wang SB, Li Y, Xu DD, Ungvari GS, Ng CH, et al. Prevalence of suicidal behaviors in patients with major depressive disorder in China: a comprehensive meta-analysis. J Affect Disord 2018 Jan 01;225:32-39. [doi: 10.1016/j.jad.2017.07.043] [Medline: 28779680]
8. Depression. World Health Organization. 2020. URL: https://www.who.int/health-topics/depression#tab=tab_1 [accessed 2022-04-16]
9. Lee YY, Stockings EA, Harris MG, Doi SA, Page IS, Davidson SK, et al. The risk of developing major depression among individuals with subthreshold depression: a systematic review and meta-analysis of longitudinal cohort studies. Psychol Med 2019 Jan;49(1):92-102. [doi: 10.1017/S0033291718000557] [Medline: 29530112]
10. Rodríguez MR, Nuevo R, Chatterji S, Ayuso-Mateos JL. Definitions and factors associated with subthreshold depressive conditions: a systematic review. BMC Psychiatry 2012 Oct 30;12:181 [FREE Full text] [doi: 10.1186/1471-244X-12-181] [Medline: 23110575]
11. 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 Dec;210(2):119-124 [FREE Full text] [doi: 10.1192/bjp.bp.116.188078] [Medline: 27908899]
12. Lee Y, Brietzke E, Cao B, Chen Y, Linnaranta O, Mansur RB, Global Alliance for Chronic Diseases (GACD) Mental Health Guidelines Working Group. Development and implementation of guidelines for the management of depression: a systematic review. Bull World Health Organ 2020 Oct 01;98(10):683-697H [FREE Full text] [doi: 10.2471/BLT.20.251405] [Medline: 33177758]
13. Ferenchick EK, Ramanuj P, Pincus HA. Depression in primary care: part 1-screening and diagnosis. BMJ 2019 Apr 08;365:l794. [doi: 10.1136/bmj.l794] [Medline: 30962184]

XSL•FO
RenderX

14. Maurer DM, Raymond TJ, Davis BN. Depression: screening and diagnosis. Am Fam Physician 2018 Oct 15;98(8):508-515 [FREE Full text] [Medline: 30277728]
15. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P. The cost-effectiveness of PHQ screening and collaborative care for depression in New York city. PLoS One 2017;12(8):e0184210 [FREE Full text] [doi: 10.1371/journal.pone.0184210] [Medline: 28859154]
16. Qin X, Wang S, Hsieh C. The prevalence of depression and depressive symptoms among adults in China: estimation based on a national household survey. China Econ Rev 2018 Oct;51:271-282. [doi: 10.1016/j.chieco.2016.04.001]
17. Guo L, Wang T, Wang W, Fan B, Xie B, Zhang H, et al. Association between habitual weekday sleep duration and depressive symptoms among Chinese adolescents: the role of mode of birth delivery. J Affect Disord 2020 Mar 15;265:583-589. [doi: 10.1016/j.jad.2019.11.095] [Medline: 31759667]
18. Guo L, Deng J, He Y, Deng X, Huang J, Huang G, et al. Prevalence and correlates of sleep disturbance and depressive symptoms among Chinese adolescents: a cross-sectional survey study. BMJ Open 2014 Jul 29;4(7):e005517 [FREE Full text] [doi: 10.1136/bmjopen-2014-005517] [Medline: 25079937]
19. Guo L, Hong L, Gao X, Zhou J, Lu C, Zhang WH. Associations between depression risk, bullying and current smoking among Chinese adolescents: modulated by gender. Psychiatry Res 2016 Mar 30;237:282-289. [doi: 10.1016/j.psychres.2016.01.027] [Medline: 26833280]
20. Xiang YT, Ng CH, Yu X, Wang G. Rethinking progress and challenges of mental health care in China. World Psychiatry 2018 Jun 24;17(2):231-232 [FREE Full text] [doi: 10.1002/wps.20500] [Medline: 29856546]
21. Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001-05: an epidemiological survey. Lancet 2009 Jun 13;373(9680):2041-2053. [doi: 10.1016/S0140-6736(09)60660-7] [Medline: 19524780]
22. Li F, He H. Assessing the accuracy of diagnostic tests. Shanghai Arch Psychiatry 2018 Jun 25;30(3):207-212 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.218052] [Medline: 30858674]
23. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, WHO World Mental Health Survey Collaborators. Treatment gap for anxiety disorders is global: results of the World Mental Health Surveys in 21 countries. Depress Anxiety 2018 Mar;35(3):195-208 [FREE Full text] [doi: 10.1002/da.22711] [Medline: 29356216]
24. Liao YH, Fan BF, Zhang HM, Guo L, Lee Y, Wang WX, et al. The impact of COVID-19 on subthreshold depressive symptoms: a longitudinal study. Epidemiol Psychiatr Sci 2021 Feb 15;30:e20 [FREE Full text] [doi: 10.1017/S2045796021000044] [Medline: 33583474]
25. Bhattacharyya O, Schull M, Shojania K, Stergiopoulos V, Naglie G, Webster F, et al. Building Bridges to integrate care (BRIDGES): incubating health service innovation across the continuum of care for patients with multiple chronic conditions. Healthc Q 2016;19(2):60-66. [doi: 10.12927/hcq.2016.24701] [Medline: 27700976]
26. Pathiravasan CH, Zhang Y, Trinquart L, Benjamin EJ, Borrelli B, McManus DD, et al. Adherence of mobile app-based surveys and comparison with traditional surveys: eCohort study. J Med Internet Res 2021 Jan 20;23(1):e24773 [FREE Full text] [doi: 10.2196/24773] [Medline: 33470944]
27. Yu X, Tam WW, Wong PT, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. Compr Psychiatry 2012 Jan;53(1):95-102. [doi: 10.1016/j.comppsych.2010.11.002] [Medline: 21193179]
28. Pellas J, Damberg M. Accuracy in detecting major depressive episodes in older adults using the Swedish versions of the GDS-15 and PHQ-9. Ups J Med Sci 2021 Oct 20;126:7848 [FREE Full text] [doi: 10.48101/ujms.v126.7848] [Medline: 34754407]
29. Nan H, Lee PH, McDowell I, Ni MY, Stewart SM, Lam TH. Depressive symptoms in people with chronic physical conditions: prevalence and risk factors in a Hong Kong community sample. BMC Psychiatry 2012 Nov 14;12:198 [FREE Full text] [doi: 10.1186/1471-244X-12-198] [Medline: 23151217]
30. 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 34-57. [Medline: 9881538]
31. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
32. Qi M, Zhou SJ, Guo ZC, Zhang LG, Min HJ, Li XM, et al. The effect of social support on mental health in Chinese adolescents during the outbreak of COVID-19. J Adolesc Health 2020 Oct;67(4):514-518 [FREE Full text] [doi: 10.1016/j.jadohealth.2020.07.001] [Medline: 32753347]
33. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
34. Bastien CH, Vallières A, Morin CM. Validation of the insomnia severity index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
35. Hou X, Fu Y. Reliability and validity of stress life event scale (SLE) [in Chinese]. For All Health (Academic Edition) 2013 May 30;7(10):1-4 [FREE Full text]

XSL•FO
RenderX

36. Wu L, Tan Y, Liu Y. Factor structure and psychometric evaluation of the Connor-Davidson resilience scale in a new employee population of China. BMC Psychiatry 2017 Feb 02;17(1):49 [FREE Full text] [doi: 10.1186/s12888-017-1219-0] [Medline: 28152997]
37. Guo L, Wang T, Wang W, Huang G, Xu Y, Lu C. Trends in health-risk behaviors among Chinese adolescents. Int J Environ Res Public Health 2019 May 29;16(11):1902 [FREE Full text] [doi: 10.3390/ijerph16111902] [Medline: 31146454]
38. International guide for monitoring alcohol consumption and related harm. World Health organization. 2000. URL: http://apps.who.int/iris/bitstream/handle/10665/66529/WHO_MSD_MSB_00.4.pdf?sequence=1 [accessed 2022-04-16]
39. Shin MH, Kweon SS, Choi JS, Rhee JA, Nam HS, Jeong SK, et al. Average volume of alcohol consumed, drinking patterns, and metabolic syndrome in older Korean adults. J Epidemiol 2013;23(2):122-131 [FREE Full text] [doi: 10.2188/jea.je20120065] [Medline: 23358208]
40. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001 Nov 01;125(1-2):279-284. [doi: 10.1016/s0166-4328(01)00297-2] [Medline: 11682119]
41. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999 Mar 30;18(6):681-694. [doi: 10.1002/(sici)1097-0258(19990330)18:6<681::aid-sim71>3.0.co;2-r] [Medline: 10204197]
42. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ 2019 Apr 09;365:l1476 [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]
43. Li H, Zheng D, Li Z, Wu Z, Feng W, Cao X, et al. Association of depressive symptoms with incident cardiovascular diseases in middle-aged and older Chinese adults. JAMA Netw Open 2019 Dec 02;2(12):e1916591 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.16591] [Medline: 31800066]
44. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. The Lancet Psychiatry 2019 Mar;6(3):211-224 [FREE Full text] [doi: 10.1016/s2215-0366(18)30511-x] [Medline: 30792114]
45. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. J Cell Mol Med 2019 Apr 07;23(4):2324-2332 [FREE Full text] [doi: 10.1111/jcmm.14170] [Medline: 30734486]
46. Uzun S, Kozumplik O, Topić R, Jakovljević M. Depressive disorders and comorbidity: somatic illness vs. side effect. Psychiatr Danub 2009 Sep;21(3):391-398 [FREE Full text] [Medline: 19794363]
47. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan 17;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
48. Vetter JS, Spiller TR, Cathomas F, Robinaugh D, Brühl A, Boeker H, et al. Sex differences in depressive symptoms and their networks in a treatment-seeking population - a cross-sectional study. J Affect Disord 2021 Jan 01;278:357-364 [FREE Full text] [doi: 10.1016/j.jad.2020.08.074] [Medline: 33002727]
49. Chiang YH, Beckstead JW, Lo SC, Yang CY. Association of auditory hallucination and anxiety symptoms with depressive symptoms in patients with schizophrenia: a three-month follow-up. Arch Psychiatr Nurs 2018 Aug;32(4):585-590. [doi: 10.1016/j.apnu.2018.03.014] [Medline: 30029752]
50. Duarte-Tagles H, Salinas-Rodríguez A, Idrovo AJ, Búrquez A, Corral-Verdugo V. [Biodiversity and depressive symptoms in Mexican adults: Exploration of beneficial environmental effects]. Biomedica 2015 Aug;35 Spec:46-57. [doi: 10.1590/S0120-41572015000500006] [Medline: 26535741]
51. Stahl ST, Beach SR, Musa D, Schulz R. Living alone and depression: the modifying role of the perceived neighborhood environment. Aging Ment Health 2017 Oct 07;21(10):1065-1071 [FREE Full text] [doi: 10.1080/13607863.2016.1191060] [Medline: 27267633]
52. Li L, Wu C, Gan Y, Qu X, Lu Z. Insomnia and the risk of depression: a meta-analysis of prospective cohort studies. BMC Psychiatry 2016 Nov 05;16(1):375 [FREE Full text] [doi: 10.1186/s12888-016-1075-3] [Medline: 27816065]
53. Badrasawi M, Zidan S. Prevalence and correlates of depressive symptoms in older people in the West Bank, Palestine: cross-sectional study. East Mediterr Health J 2021 Mar 23;27(3):260-268 [FREE Full text] [doi: 10.26719/2021.27.3.260] [Medline: 33788215]
54. Caouette JD, Guyer AE. Cognitive distortions mediate depression and affective response to social acceptance and rejection. J Affect Disord 2016 Jan 15;190:792-799 [FREE Full text] [doi: 10.1016/j.jad.2015.11.015] [Medline: 26625091]
55. Coryell W, Fiedorowicz JG, Solomon D, Leon AC, Rice JP, Keller MB. Effects of anxiety on the long-term course of depressive disorders. Br J Psychiatry 2012 Mar;200(3):210-215 [FREE Full text] [doi: 10.1192/bjp.bp.110.081992] [Medline: 21984801]
56. Jackson ML, Sztendur EM, Diamond NT, Byles JE, Bruck D. Sleep difficulties and the development of depression and anxiety: a longitudinal study of young Australian women. Arch Womens Ment Health 2014 Jun;17(3):189-198. [doi: 10.1007/s00737-014-0417-8] [Medline: 24647705]
57. Okun ML, Luther JF, Wisniewski SR, Wisner KL. Disturbed sleep and inflammatory cytokines in depressed and nondepressed pregnant women: an exploratory analysis of pregnancy outcomes. Psychosom Med 2013 Sep;75(7):670-681 [FREE Full text] [doi: 10.1097/PSY.0b013e31829cc3e7] [Medline: 23864582]

XSL•FO
RenderX

58. Chen L, Wang L, Qiu XH, Yang XX, Qiao ZX, Yang YJ, et al. Depression among Chinese university students: prevalence and socio-demographic correlates. PLoS One 2013;8(3):e58379 [FREE Full text] [doi: 10.1371/journal.pone.0058379] [Medline: 23516468]
59. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
60. Kang HJ, Kim SY, Bae KY, Kim SW, Shin IS, Yoon JS, et al. Comorbidity of depression with physical disorders: research and clinical implications. Chonnam Med J 2015 Apr;51(1):8-18 [FREE Full text] [doi: 10.4068/cmj.2015.51.1.8] [Medline: 25914875]
61. Hettema JM. What is the genetic relationship between anxiety and depression? Am J Med Genet C Semin Med Genet 2008 May 15;148C(2):140-146. [doi: 10.1002/ajmg.c.30171] [Medline: 18412101]
62. Grant F, Guille C, Sen S. Well-being and the risk of depression under stress. PLoS One 2013 Jul 01;8(7):e67395 [FREE Full text] [doi: 10.1371/journal.pone.0067395] [Medline: 23840872]
63. Keyes CL. Mental illness and/or mental health? Investigating axioms of the complete state model of health. J Consult Clin Psychol 2005 Jun;73(3):539-548. [doi: 10.1037/0022-006X.73.3.539] [Medline: 15982151]
64. Wood AM, Joseph S. The absence of positive psychological (eudemonic) well-being as a risk factor for depression: a ten year cohort study. J Affect Disord 2010 May;122(3):213-217. [doi: 10.1016/j.jad.2009.06.032] [Medline: 19706357]
65. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]

Abbreviations

BRIDGES: Building Bridges to Integrate Care
CD-RISC: Connor-Davidson Resilience Scale
DCC: Depression Cohort in China
GP: general practitioner
HR: hazard ratio
LMIC: low- and middle-income countries
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatric Interview
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
SLESQ: Stressful Life Events Screening Questionnaire
WHO: World Health Organization
WHO-5: 5-item World Health Organization Well-Being Index

Edited by T Leung; submitted 03.06.21; peer-reviewed by T Tillmann, L Costantini, G Mayer; comments to author 14.01.22; revised version received 18.02.22; accepted 30.03.22; published 20.05.22.

Please cite as:
Zhang H, Liao Y, Han X, Fan B, Liu Y, Lui LMW, Lee Y, Subramaniapillai M, Li L, Guo L, Lu C, McIntyre RS
Screening Depressive Symptoms and Incident Major Depressive Disorder Among Chinese Community Residents Using a Mobile App–Based Integrated Mental Health Care Model: Cohort Study
J Med Internet Res 2022;24(5):e30907
URL: https://www.jmir.org/2022/5/e30907
doi:10.2196/30907
PMID:35594137

©Huimin Zhang, Yuhua Liao, Xue Han, Beifang Fan, Yifeng Liu, Leanna M W Lui, Yena Lee, Mehala Subramaniapillai, Lingjiang Li, Lan Guo, Ciyong Lu, Roger S McIntyre. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Use of Social Media as a Persuasive Platform to Facilitate Nutrition and Health Behavior Change in Young Adults: Web-Based Conversation Study

Vanessa J Friedman[1,2], BBiomedSci (Hons); Cassandra J C Wright[1,2,3], PhD; Annika Molenaar[2], BNutSci (Hons); Tracy McCaffrey[2], PhD; Linda Brennan[4], PhD; Megan S C Lim[1,2,5], PhD

[1]Burnet Institute, Melbourne, Australia
[2]Monash University, Melbourne, Australia
[3]Menzies School of Health Research, Darwin, Australia
[4]Royal Melbourne Institute of Technology, Melbourne, Australia
[5]University of Melbourne, Carlton, Australia

Corresponding Author:
Megan S C Lim, PhD
Burnet Institute
85 Commercial Rd
Melbourne, 3004
Australia
Phone: 61 85062403
Email: megan.lim@burnet.edu.au

Abstract

Background: Globally, suboptimal dietary choices are a leading cause of noncommunicable diseases. Evidence for effective interventions to address these behaviors, particularly in young adults, is limited. Given the substantial time young adults spend in using social media, there is interest in understanding the current and potential role of these platforms in shaping dietary behavior.

Objective: This study aims to explore the influence of social media on young adults' dietary behaviors.

Methods: We recruited 234 young adults aged 18-24 years and living in Australia, using market and social research panels. We applied a digital ethnography approach to collect data from web-based conversations in a series of forums, where participants responded to different health-themed questions related to health behavior change and persuasion on social media. We conducted a qualitative thematic analysis.

Results: Participants described how social media influenced their decisions to change their health behaviors. Access to social support and health information through web-based communities was juxtaposed with exposure to highly persuasive fast-food advertisements. Some participants expressed that exposure to web-based health-focused content induced feelings of guilt about their behavior, which was more prominent among women. Fast-food advertisements were discussed as a contributor to poor health behaviors and indicated as a major barrier to change.

Conclusions: Young adults reported that social media is highly persuasive toward dietary behavior through different pathways of social influence. This suggests that social norms on the web are an important aspect of changing young adults' health behaviors. The commercialization of social media also encourages poor health behaviors, largely through fast-food advertisements. Future social media–delivered dietary interventions should acknowledge the social and environmental factors that challenge the ability of young adults to make individual health behavior improvements. Care should also be taken to ensure that future interventions do not further elicit guilt in a way that contributes to poor mental health within this community.

(J Med Internet Res 2022;24(5):e28063) doi:10.2196/28063

KEYWORDS

young adults; nutrition; physical activity; mental health; social media; qualitative methods; health promotion

XSL•FO
RenderX

Introduction

Young Adults' Health and Nutrition

The prevalence of noncommunicable diseases is increasing globally, and they have become a leading health concern. It is known that suboptimal diets with low intake of fruits and vegetables and high intake of processed fatty foods are contributing to this trend [1]. Therefore, it is concerning that the diet quality of young adults typically reduces as they navigate the challenging shift from adolescence to adulthood [2]. Transitioning away from school and family resources toward workforce or further study has a strong and lasting impact on a young adult's dietary behaviors [3,4]. Studies show that many young adults prioritize other aspects of their lives over healthy eating, which may be perceived as expensive and time-consuming for an age group that typically has low income [5,6]. The transitional nature of young adulthood can also present challenges in creating targeted and effective dietary interventions to reach this population [4]. Previous research has shown that young adults conceptualize health more broadly than physical health or the prevention of chronic disease and value mental, social, financial, and spiritual aspects [7]. As such, to develop engaging, feasible, and acceptable approaches that target dietary behaviors in this age group, the focus needs to extend beyond long-term health and should incorporate holistic views and short-term benefits [7].

Previous qualitative studies suggest that dietary behaviors during young adulthood are strongly influenced by internal perceptions and social norms [5,8,9]. Young adults were found to base their food choices on what they perceived their friends or family were eating, highlighting the power of socially normative messages in this domain [9]. Social media has become a key component of the social environment of young adults [8,10,11]. The ability to share, comment on, and react to other users' posts increases interactions in this setting [12]. Social media delivers a constant stream of social input to young adults and has become a place for them to view and compare themselves with idealized versions of both their peers' and strangers' lives [7]. A recent systematic review indicated that image-related comparisons on social media may negatively impact the body image of young adults and drive poor eating behaviors such as restriction or overeating [13].

Social Media and Nutrition

In 2018, a total of 99% of Australians aged 18-29 years used social media regularly, with 89% of them accessing their accounts at least once daily [14]. High rates of social media use have led food brands and companies to use social media to enhance their engagement with young adults [15]. Many fast-food companies use largely unregulated social media advertising regimes to promote energy-dense nutrient-poor foods that are shared throughout young adult peer networks [8,15]. Social media *influencers* have emerged as key players in these marketing strategies [16,17]. They are recognized as people who hold persuasive power by sharing their lives on various platforms and forming emotional connections with their audiences [16,17]. As such, companies work with influencers who provide paid product reviews to their audiences to boost the company's sales and consumer engagement rates [17]. Some influencers exclusively post health and lifestyle content; however, many of these *health-focused influencers* lack professional accreditation and may post misleading nutrition advice that is not evidence-based [18]. Currently, experts in nutrition are becoming less trusted [19], and social media users are more likely to engage with and trust health-focused content from influencers than that from food industry or health promotion [20]. Hence, commentary from influencers has a relatively large impact on the values, beliefs, and behaviors of consumers regarding nutrition in both positive and negative ways [18].

The ability of social media to influence young adult audiences has also sparked interest from public health practitioners as a potential platform for health promotion [21]. Social media has previously been shown to influence health knowledge, with some studies identifying its positive influence on young adults through access to healthy recipes and exercises [10,11,22]. A recent systematic review identified that social media–delivered nutrition interventions that target adolescents and young adults lead to significant dietary improvements in 11 of 16 studies [11]. However, many of the interventions used were complex, with social media often being part of a secondary component, thus making it difficult to distinguish the true impact [11]. Our systematic review evaluated the efficacy of social media–delivered nutrition interventions in young adults only and identified that engagement with social media content varied greatly between 3% and 69% [22]. Young adults preferred to use social media passively, in a unilateral interaction, receiving information rather than sharing information [22]. Young adults were also not comfortable with talking about their weight on the web, highlighting the need to avoid weight-centric narrative in health promotion [22]. In 2 separate studies, we also found that social media users engage more frequently with food industry and lifestyle brands than with health promotion [20,23]. This highlights the need to develop more effective social media–delivered health promotion tools to encourage healthy behaviors in Australian young adults. Thus, the impact of social media on dietary behaviors must be further explored, and its persuasive abilities must be further understood.

This Study

Phase 1a of the Communicating Health project seeks to gain insight into the use and application of social media, as it relates to 12 health-related and eating-related topics. This will allow for the identification of the channels, tones, and content-types that have the greatest potential for health promotion development. To understand how to develop effective social media–delivered health promotion tools, first, it is important to explore whether social media is currently impacting young adult dietary behaviors, and if so, how it is being used as a platform for persuasion. As such, this study aims to investigate what prompts young adults to make positive health and nutrition behavior changes and to understand the influence of social media as a persuasive medium on young adults' health and nutrition behaviors.

XSL•FO
RenderX

Methods

Web-Based Conversations

This study is part of the larger Communicating Health project [24], which is a multistage mixed methods study that explores the dietary behaviors and social media use of Australian young adults. It aims to develop health promotion strategies using social marketing techniques. The data used in this study form a part of the formative phase of the Communicating Health project, phase 1a, which involved web-based conversations that explored young adults' health, eating behaviors, and social media use [24]. An outline of all 4 phases of the Communicating Health project has been published previously [24]. The qualitative web-based conversations were hosted by an independent market research field house over a 4-week period. The web-based conversations were prompted by questions posed by the market research facilitators in moderated and secure web-based chat rooms. This method is based on digital ethnographic principles to understand how the digital aspects of society interact with the other material, sensory, and social aspects of human existence [25,26]. As a responsive data collection technique commonly used in consumer behavior research, web-based conversations allowed participants to interact with each other for a longer period than that allowed in traditional focus groups or interviews [27].

Ethics Approval and Consent to Participate

Phase 1a received ethics approval from the RMIT Business College Human Ethics Advisory Network (project number 20489) and the Monash University Human Research Ethics Committee (project number 7807). Participants consented to anonymized findings being published when they completed the patient information and consent form before participating in the study. Ethics approval for this project was granted by the Monash University Human Research Ethics Committee (project ID 19417).

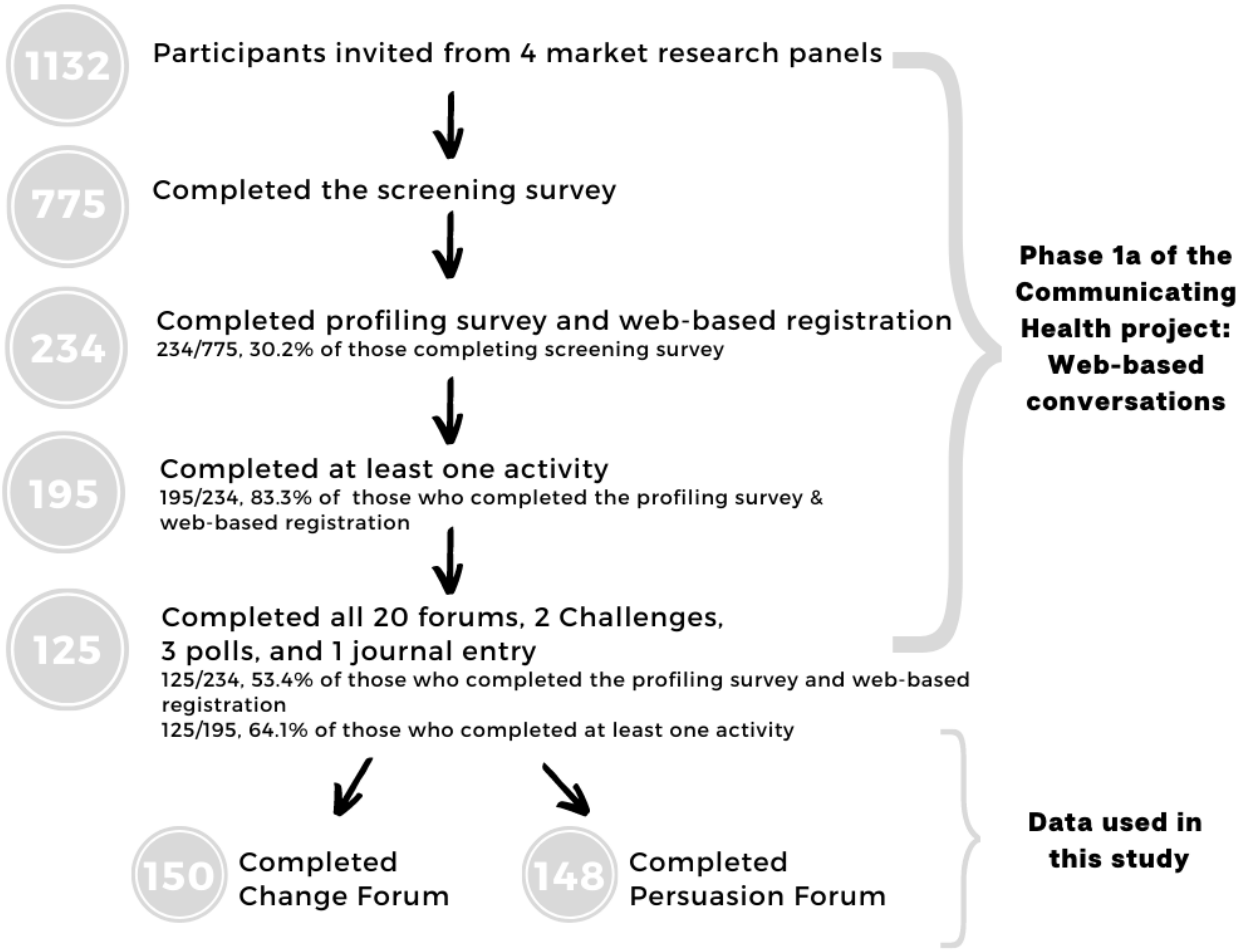
Recruitment

Guided by previous studies with similar methodologies [28], the recruitment target was 200 young adults aged 18-24 years, living in Australia, and using social media at least twice a day. The recruitment period was from May 2017 to June 2017. This process was facilitated by an Australian Research Society–certified field house [29]. Young adults who had previously provided consent to participate in the research by signing up to market research panels were invited to participate in this study. Participants were from 3 research panels that were accredited by the International Organization for Standardization [30-32].

Panel members were sent an email invitation to complete a short screening questionnaire to assess their eligibility. Then, those who were eligible were asked to complete a profiling survey for collecting demographic information, self-reported weight and height (to calculate BMI [kg/m^2]), social media use, and interest in health. Health interest (low or moderate and high) was determined by the median value of the following question asked in the profiling survey: "On a scale of 1-7 where 1 means 'Strongly Disagree' and 7 means 'Strongly Agree', please indicate how strongly you agree with the following statement - I take an active interest in my health." The profiling survey was completed by 234 participants, who were then provided a link through email to sign up to the web-based conversation website. The participant flow diagram is shown in Figure 1. Then, the participants were stratified into 4 communities based on their age (18-21 years and 22-24 years) and interest in health (low or moderate and high). Those with low interest in health were grouped separately from those with moderate and high interest in health and, then, further divided by age, leading to 4 groups with 42-60 participants each. Profiling was set to achieve approximately equal number of participants in all groups and an approximately nationally representative distribution of gender and location (both Australian State or Territory and location type; ie, metropolitan and regional locations) [33]. All 4 communities had access to the same forums but could only interact with the members within their community. The dropout rate was high, which was expected for this age group. Therefore, a referral system was established, in which existing participants could refer a friend, who was then screened and profiled in the same way.

XSL•FO
RenderX

Figure 1. Participant flow diagram.



Data Collection

The web-based conversations were conducted on a private web-based forum portal from May 10, 2017, to June 6, 2017. There were 20 forums in total (each took approximately 5 minutes to complete), 3 short polls, and an ongoing journal entry to which the participants were asked to contribute at least four times. The participants were also asked to complete 2 different challenges. The first challenge invited the participants to come up with a creative way to make more young adults to eat more fruits and vegetables. The second challenge asked the participants to persuade someone to kick-start a healthy lifestyle, using ≤160 characters. Then, the participants were able to see each other's ideas and comment on their favorites. These different aspects of the web-based conversations explored different areas of health, nutrition, and social media, and participants responded to prompts from the market research moderators. The forums were released at different times but remained open for the 4-week period. Owing to different aspects being released on separate weeks, there were different numbers of participants who completed each forum (Figure 1). As per the standard practice by Australian Market and Social Research Society Limited, the participants were reimbursed for their time with a gift voucher worth Aus $100 (US $74.9) upon completion of all aspects of the web-based conversations, with a further Aus $100 (US $74.9) given to the 5 most descriptive and detailed forum responses from each web-based community (ie, 20 in total). This study reports on 2 of the forums that formed the web-based conversations: *catalysts for change* (referred to as *change*) and *persuasion on social media* (referred to as *persuasion*). These were chosen for analysis because they discussed health behavior change and persuasion on social media, which aligned with the research aim of this study. Table 1 describes the prompt questions used in both forums.

XSL•FO
RenderX

Table 1. Forum prompt questions.

Forum title	Discussion guide	Logic of enquiry
Catalysts for change	• Have you changed anything recently to make you happier? Healthier? • What triggered that change? • How did you go about making that change? • Did it change the dynamic within your friendship circle? • Have you kept up with that change? (for how long – or probe for that?) What (if anything) was pulling you back to your old ways? • Did social media give you any inspiration? Help? Hinder? Did anything else or any other tools play a role too (eg, apps, websites or even just people...)?	An exploration of what prompted lasting health behavior change in the young adult participants and whether social media played a role in this process.
Persuasion on social media	• Can you think of times when you have used social media to persuade others to do something? • Can you think of times when you have been persuaded? • More broadly, can you think of how social media has influenced things you do in relation to health and healthy lifestyle?	To determine whether the participants viewed social media as a persuasive medium and to understand some of the factors that influenced its persuasiveness.

Data Analysis

Descriptive analyses of the profiling survey data were conducted (Table 2). We undertook a thematic analysis of the forum responses using open coding and inductive techniques in alignment with a constructivist interpretation. As detailed in Table 3, this was guided by the systematic 6-step approach by Braun and Clarke [34]: data familiarization, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and, finally, producing the report. This process enabled the large data set to be broken down into small units of comparison (codes) and rebuilt into themes that provided a systematic description of the participants' experiences of health behavior change and social media use. A constant comparison approach was used to identify themes associated with healthy eating and other health behaviors [35]. Within this approach, each data point was compared with pre-existing data points to identify commonalities that existed within and between the participants' responses. These commonalities became codes, and then, similar codes were grouped into broader themes that encapsulated the data set. Investigator triangulation was used to reduce subjectivity and enhance the rigor of the research findings [36]. As such, double-blind coding was conducted by 2 authors (VF and AM), who both coded each data set independently and, then, met to discuss their findings and reach an agreement. The researcher and primary coder of this study (VF) had a background in biomedical science and belonged to the age group of the participants. Growing up with social media, her empathetic connection to the study allowed her to extract details that may have otherwise been missed. The other coder (AM) had a disciplinary background in nutrition, which may have influenced her perception of the participants' responses. The different backgrounds of each coder allowed the data to be examined from 2 different perspectives, which decreased subjectivity in the interpretation of the results. Although the forums occurred separately, data from both forums were analyzed together as overlapping themes were identified. Owing to different prompts in each forum (Table 1), some forums contributed more to certain themes than others.

XSL•FO
RenderX

Table 2. Demographic information of the participants who completed the change forum (n=150[a]).

Variable and category	Participants, n (%)
Gender identity	
Women	87 (58)
Men	62 (41.3)
Nonbinary, gender-fluid, or genderqueer	1 (0.7)
Age (years); health interest level (low or moderate and high)	
18-21; low	38 (25.3)
18-21; moderate and high	47 (31.3)
22-24; low	31 (20.7)
22-24; moderate and high	34 (22.7)
State	
Australian Capital Territory	1 (0.7)
New South Wales	45 (30)
Northern Territory	1 (0.7)
Queensland	21 (14)
South Australia	10 (6.7)
Tasmania	4 (2.6)
Victoria	42 (28)
Western Australia	26 (17.3)
Language spoken at home	
Culturally and linguistically diverse	42 (28)
English	108 (72)
BMI (kg/m^2)	
<18.5 (underweight)	16 (10.7)
18.5-24.9 (healthy weight)	82 (54.7)
25-29.9 (overweight)	33 (22)
≥30 (obese)	19 (12.7)
Currently studying	
No	47 (31.3)
Yes	103 (68.7)
Current level of study	
High school (year 12)	7 (4.7)
Technical and further education, college, or diploma	12 (8)
University undergraduate course	75 (50)
University postgraduate course	9 (6)

[a]Of the 150 participants who completed the change forum, 148 (98.7%) participants completed the persuasion forum.

XSL•FO
RenderX

Table 3. Approach to thematic analysis [34].

Step	Outline	How it applied to the analysis of both forums: catalysts for change (change) and persuasion on social media (persuasion)
1—Data familiarization	Full read through of the data set and noting emerging concepts and codes.	• Each forum was read through twice. • Emerging patterns and potential codes were noted for each.

2—Generating initial codes	Systematic identification and coding of relevant phenomena to generate a long list of codes.	• A total of 3 rounds of coding were conducted for each forum.
		• Round 1
		  • Inductive line-by-line coding: each line of the data set was labeled based on its content, identifying novel, and expected codes.
		  • Similar codes were collapsed and redundant codes (ie, scarcely present in the data set) were deleted, resulting in 57 codes for the change forum and 76 codes for the persuasion forum.
		• Round 2
		  • Key elements were focused on (change: drivers of health behavior change; persuasion: how social media acted persuasively); emerging themes were noted.
		  • Codes were further collapsed, resulting in 54 codes for the change forum and 72 codes for the persuasion forum.
		  • VF and AM met to discuss the change forum—high level of agreement; more explicit coding was needed for health behaviors beyond nutrition (exercise, smoking, and alcohol) in VF's codes.
		• Round 3
		  • Codes in the change forum were drawn out more distinctly, and coding for the persuasion forum focused on the impact of persuasion on health behaviors; emerging themes were noted.
		  • VF and AM met to discuss the persuasion forum and had complete agreement.
		  • Final collapsing and deleting resulted in 61 codes in each forum.

3—Searching for themes	Codes are compared and grouped into common themes. Considering the relationships between codes, emerging subthemes are generated.	• Relationships between codes within each forum were considered to develop appropriate themes. Theme maps were generated in a hand-written format, using sticky notes to move codes around until they fit in the most logical sequence. These candidate themes were approved by AM with some slight adjustments to be made in the change forum.

4—Reviewing themes	All themes are reviewed for their relevance to the overall data set. Themes must be internally homogenous (contain similar codes) and externally heterogeneous (each theme is distinct).	• Candidate themes were assessed for coherence (internal homogeneity) and distinction (external heterogeneity) in 2 steps:
		  • All extracts within each code were reviewed to ensure that they fit logically and were coherent.
		  • The data set was reread against the theme map to ensure that the themes were valid and representative of the overall data set. Any previously missed extracts were coded: change forum (2 new codes) and persuasion forum (0 new codes).

5—Defining and naming themes	The essence of what a theme captures within the data set is formed with textual evidence identified to portray each point. Each theme has a clear scope and succinct heading.	• All themes were revised with textual evidence obtained from the data set to support each included code. Dot-point summaries were written for each theme and subtheme to capture their meaning, which were later developed into more comprehensive outlines. Theme and subtheme headings were developed. Throughout the process, a large degree of overlap emerged between the change and persuasion forums. As such, the themes were readjusted to combine the 2 analyses.
6—Producing the written report	The themes are collated into a written report that provides a description of the data set using extracts from the data set.	• An integrated analysis of the change and persuasion forums was synthesized by incorporating textual evidence and written explanations to present the findings of the data set.

XSL•FO
RenderX

Results

Participant Characteristics

The characteristics of the participants based on self-reported data are presented in Table 2. Each forum had a different completion rate; *change* was completed by 76.9% (150/195), and *persuasion* was completed by 75.9% (148/195). Of the 150 respondents, most were women (n=87, 58%), reported moderate or high interest in health (n=81, 54%), had a healthy BMI (n=82, 54.7%), and were currently studying (n=103, 68.7%). Most participants lived in New South Wales (45/150, 30%) or Victoria (42/150, 28%) and spoke English at home (108/150, 72%).

Thematic Analysis

Overview

Following a manual thematic analysis of both forums, the data sets were collated to develop 4 major themes from key recurring concepts. These included (1) peer support, (2) access to influencers and web-based communities, (3) advertising, and (4) constant exposure to content (Textbox 1).

Textbox 1. A brief description of the major themes from the data set.

Theme 1: peer support

- Many participants described that social media provided them with access to support from their *real-world* peers who helped them to make and maintain a health behavior change. Alternatively, peer influence sometimes led to negative health behaviors such as buying fast food. Overall, participants simply wanted to be involved in what their peers were doing, whether this meant attending an exercise class together or buying fast food.

Theme 2: access to influencers and web-based communities

- Health-focused communities managed by persuasive social media influencers were considered by many participants as sources of support for making positive health behavior changes. Some participants believed that these communities enhanced their willpower, whereas other participants shared that they influenced their health attitudes but not their health behaviors. These communities also tended to promote an image-based perspective of health, which evoked feelings of guilt in some participants.

Theme 3: advertising

- Participants described social media newsfeeds that were flooded with advertisements, which they found persuasive. This included health products; however, fast-food advertisements had a more dominant presence. Promotions based on taste and affordability prompted fast-food purchases, which some participants viewed as lack of willpower.

Theme 4: constant exposure to content

- The design of social media to constantly expose its users to content was described as highly persuasive. Repeated exposure to health-themed content made the participants more conscious of their own health behaviors. Some participants explained that constantly viewing health content alongside fast-food advertisements made them feel conflicted and guilty if they consumed fast food.

Peer Support

The young adults in our study noted that their peers influenced their health behaviors through both direct communication on the web and exposure to the content they posted on social media feeds. A participant shared the following:

> *One of my friends would always message me to double check I was having breakfast and would always ask me what I had because she could tell when I was lying.* [change forum; female; aged 18-21 years; low interest in health]

Participants reported feeling influenced to make health behavior improvements when their peers posted on social media about their own positive changes. This was exemplified by a participant who wrote the following:

> *I also sometimes get influenced to be fitter when people share on social media their own health transformations. I then reconsider my lifestyle and see what I can do to be more active and healthy myself.* [persuasion forum; male; aged 18-21 years; low interest in health]

Many participants shared that peer influence on social media drove both positive and negative health behaviors based on the content being shared:

> *Knowing what my friends eat...can encourage me to eat certain things. When a person shares an exercise [post], I am more inclined myself to exercise...when a friend makes a comment on a [Fast food brand name removed]'s post, I am more inclined to check out their deals.* [persuasion forum; male; aged 22-24 years; moderate and high interest in health]

Some participants, most of whom were men, also shared their health activities with peers on social media to reach out for support and hold themselves accountable. A participant explained the following:

> *To put myself out there on social media really gave me the confidence and gumption to stick to my routine.* [persuasion forum; male; aged 22-24 years; low interest in health]

Access to Influencers and Web-Based Communities

In addition to receiving peer support from friends or acquaintances, the participants also experienced community support through health-focused pages or communities on social

XSL•FO
RenderX

media, which were often managed by health-focused influencers. These communities provided participants with access to useful diet and lifestyle information and motivated them by providing a sense of unity and connection with others. A participant shared the following:

> *Social media has had an extremely positive influence on me when it comes to maintaining a healthy lifestyle...It's 24/7 access to help, reassurance and motivation.* [persuasion forum; female; aged 22-24 years; moderate and high interest in health]

Female participants with moderate and high interest in health more commonly discussed social media as a positive influence.

The participants also described willpower as an important moderator of the influence that social media content had on their behavior. A participant believed the following:

> *If I try hard enough to work on my eating and doing more exercise then I will be able to be like them [models] with their hundreds of likes on their photos.* [persuasion forum; male; aged 18-21 years; moderate and high interest in health]

Web-based health communities could motivate these highly driven participants to remain self-disciplined and self-reliant, which helped them to resist negative external influences such as fast-food advertisements. A participant explained the following:

> *I find it's easier to stay motivated if I stay home and in routine (without access to bad food of course), and interact regularly with the online fitness/health community.* [change forum; female; aged 22-24 years; moderate and high interest in health]

This was most often discussed by participants aged 18-21 years, with moderate and high interest in health. In contrast, some participants revealed that health-focused communities influenced their attitudes toward health, but not always their behaviors. A participant shared the following:

> *I'm following many health and 'fitspo' blogs and pages which teach me simple recipes and exercise regimes- now whether I actually follow them or not is another question.* [persuasion forum; female; aged 18-21 years; low interest in health]

This concept was most often described by women aged 18-21 years, with low interest in health.

A disadvantage of social media health-focused communities described by some participants was their tendency to portray health in an "image-fueled way" (*persuasion* forum; female; aged 18-21 years; moderate and high interest in health). For some participants, this led to feelings of self-doubt from upward comparisons with photos of others on the web. This was exemplified by a participant who wrote the following:

> *Seeing health/fit looking people on social media...can either inspire people to be healthier or they can discourage people as their body/lifestyle/look is unattainable.* [persuasion forum; female; aged 18-21 years; moderate and high interest in health]

Participants aged 18-21 years more readily associated guilt with health content on social media. The female participants were largely discouraged by this guilt, whereas the male participants discussed that upward comparisons motivated them to make a change.

Other participants described an awareness that their repeated exposure to social media health-focused influencers affected their outlook on which health behaviors they adopted:

> *While I love hiking and outdoors activities i'm not sure whether that's entirely due to my own interests or because I see social media influencers with the perfect life doing things like that too...maybe I feel like that's what having a healthy balanced life is like because that's how it's portrayed on social media.* [persuasion forum; female; aged 22-24 years; low interest in health]

Some participants also expressed indifference or disinterest toward social media in general. A participant stated the following:

> *Social media hasn't really changed anything, because I don't really like sharing my thoughts and activities through social media.* [change forum; male; aged 22-24 years; moderate and high interest in health]

This viewpoint was more commonly described by male participants. Similarly, most participants did not engage with social media as a means to actively interact with others, but used it as a purely observational platform. This was most commonly observed in those participants who were aged 22-24 years, with moderate or high interest in health.

Advertising

Participants reported that advertisements on social media also had an impact on their drive to make a health behavior change. Many participants described being drawn to health products that advertised benefits such as weight loss, feeling better, or affordability. A participant was allured by a detox tea, as it claimed "to help prevent bloating, slim your tummy and give you extra vitamins" (*persuasion* forum; female; aged 18-21 years; low interest in health). Participants aged 18-21 years, with low interest in health, more often discussed the persuasive effects of social media advertisements. Exposure to advertisements from large corporations that appeared *randomly* on participants' newsfeeds were more often discussed than paid influencer content or products.

Fast-food advertisements were described as having a dominant presence on social media newsfeeds, which influenced the participants' food choices. A participant explained the following:

> *Most ads on Facebook influence my health negatively...as they are usually for unhealthy food options.* [persuasion forum; male; aged 22-24 years; low interest in health]

This notion was discussed more often by female participants and those aged 18-21 years. Male participants more commonly referred to cost-based advertisements of fast food as persuasive:

XSL•FO
RenderX

> *These [fast-food] meals are cheap and easy, [and] although they're[sic] aren't healthy I know they will taste good. This [fast-food] advertising is very persuasive as it makes me believe that I am hungry and I can not[sic] stop thinking about the new promotion.* [persuasion forum; male; aged 18-21 years; moderate and high interest in health]

Some participants who revealed the difficulties in resisting fast-food advertisements viewed their temptation as a lapse in self-discipline. A participant remarked the following:

> *I may have been 'persuaded' (read 'reminded of my weak will') to purchase [fast food brand name removed] on several occasions.* [persuasion forum; male; aged 18-21 years; low interest in health]

Constant Exposure to Content

The participants described that being frequently exposed to health-themed or food-themed content was a highly influential aspect of social media. They believed that the way content was presented on social media was more persuasive than the content itself, explaining that they were more likely to engage with something if it appeared frequently in their newsfeeds. This was exemplified by a participant who shared the following:

> *While social media can be used as a platform...to persuade, I really thing [sic] social media...does most of the persuading [itself].* [persuasion forum; female; aged 18-21 years; moderate and high interest in health]

Although this was discussed in reference to fast-food advertisements, it was more often applied to health content, which encouraged some participants to be more health-conscious:

> *I see a lot about healthy lifestyle and fitness in my social media feeds and I think that constant exposure has made me much more conscious of the choice I make, and a bit more aware of exercising and eating healthy.* [persuasion forum; female; aged 22-24 years; low interest in health]

Although constant exposure to both general and health-themed content increased engagement in healthy behaviors in most participants, a participant described that they "didn't want to engage in something that was being shoved in my face every time I opened Facebook, Twitter, and even Instagram" (persuasion forum; female; aged 18-21 years; low interest in health).

Some participants described that the cohabitation of health content and food temptation on the web made them feel guilty about their health behaviors. Their repeated exposure to these conflicting health ideals placed side-by-side evoked poor mental health and body image:

> *In relation to health and lifestyle it [social media] has not at all helped because it always shows videos of tasty unhealthy recipes and ads for [Fast food brand name removed] and [Fast food brand name removed]...It also then shows me photos of tall, tan, skinny models which makes me feel so bad about eating all the fast food.* [persuasion forum; female; aged 18-21 years; low interest in health]

Discussion

Principal Findings

This study aimed to contribute to a growing body of research that defines the interplay between young adults' health behaviors and social media. Specifically, this study aimed to address current gaps regarding what prompts young adults to make positive health and nutrition behavior changes and to understand how social media acts as a platform for persuasion in this process. Our study found that peer support was crucial in shaping young adults' health behaviors and that using social media to both message friends and view their posts prompted change. Web-based health-focused communities were also identified as a source of support, and health-focused influencers at their helm were found to play a prominent role in persuasion. Other persuasive aspects of social media included fast-food advertisements and constant exposure to content through newsfeeds. These aspects influenced participants' health behaviors, particularly regarding purchasing fast food or being more conscious of dietary choices.

The participants of this study highlighted social influence as a key driver of health behavior change in social media–based peer interactions. The significance of real-world peer influence on young adults' health behavior change is well established in the literature [37-39]. The results of this study suggest that young adults also find valuable social support in web-based environments. For example, seeing peers posting about their own healthy behaviors inspired some participants to follow their lead. In contrast, participants were also persuaded to purchase fast food if their friends were sharing posts from these brands. As such, this study indicates that young adults are likely to align their health behaviors with the actions of their peers, regardless of whether it is a positive or negative action. These findings are supported by social cognitive theory, which posits that people will mimic their peers to gain social acceptance [40]. Moving forward, targeting peer networks rather than individuals may enhance social media–delivered health promotion techniques. For this to be done effectively, further research may be needed to gain a greater understanding of how peer networks communicate on social media.

In addition to peers, the participants also identified health-focused influencers and web-based communities as having persuasive power over their dietary behaviors. Past research indicates that lifestyle brands, including influencer pages, on Facebook and Instagram have higher levels of engagement than both food industry and health promotion pages [20,23]. Their engagement is likely increased by their use of relatable content, positive emotional messages, paid promotions, and simple diet and exercise tricks that promise happiness by achieving appearance-related goals [20,23,41]. Our results moderately support this narrative, as some participants discussed influencers, particularly health-focused influencers, as a source of motivation to make and maintain positive health behavior change. However, some participants also explained that content from health-focused influencers only altered their attitudes

XSL•FO
RenderX

toward health and did not lead to tangible behavior changes. Moreover, advertisements from large companies were discussed more often as a source of persuasion than influencers or influencer-promoted products in this study. These issues have recently been exemplified in the *Girls Make Your Move* campaign, which received funding from the Australian Department of Health to increase the involvement of girls aged 12-21 years in sport [42]. Although influencers were involved in the social media promotion of this campaign, other techniques such as viewing advertisements on YouTube or interacting with campaign posts on social media platforms led to more tangible behavior changes [42]. Moreover, the Australian Federal Health Minister recently launched an investigation into the campaign after learning that some of the influencers involved were also sponsored by alcohol brands and displayed racist or homophobic behavior on the web [43]. Moving forward, public health organizations need to remain cautious about engaging with influencers on social media for health promotion. Furthermore, our study suggests that additional research may be needed to determine the extent of influence caused by influencers, regarding young adults' health behavior change, particularly when competing with mainstream brands for attention.

Social media environments have become heavily commercialized, and many companies pay for greater exposure to maximize their reach among young adults [44,45]. Advertisements delivered on social media is poorly regulated compared with traditional advertisements, making young users increasingly vulnerable to the persuasive tactics used by large corporations [46]. The participants of this study discussed that viewing fast-food advertisements on social media often led them to purchase fast food. The social media newsfeeds are designed such that the participants were constantly exposed to this content, which they found to be a key aspect of their persuasive abilities. Research has demonstrated that passively receiving advertisements on social media increases brand engagement and product sales, even if consumers did not have interest in the product initially [8,47]. Our study also showed that exposure to health-focused content in this manner led participants to be more aware of their health behaviors. It could be deduced that it was not the content that each individual was viewing that was persuasive, but their repeated exposure to it. The content that an individual views on their social media feeds is curated by an algorithm that predicts their likes, interests, and needs based on their behavior on the web [48], which leads to the creation of echo chambers [49]. Consequently, the more often an individual or their peers engage with social media–delivered fast-food advertisements, the more often they will be shown this content. In contrast, if an individual engages with health-focused content more regularly, this content will be displayed for them more often, which can lead to more positive health behaviors. A key issue for future social media–delivered health promotion to overcome will be ensuring that individuals with low interest in health also receive important health information that otherwise may not be *selected* for them in their echo chambers, owing to their patterns of behavior on the web.

Regardless of the heavily commercialized and persuasive setting of social media, another finding of this study was that the participants still viewed their health behaviors as an individual responsibility. This was encapsulated in the participants' beliefs that their ability to achieve a healthy lifestyle as shown to them by health-focused influencers was solely dependent on their work ethics and willpower. As described by a participant, "giving in to the temptation" of fast-food advertisements was viewed as an indicator of their own weak will, rather than the persuasive tactics used by the fast-food brand. This outlook is well documented in the literature and indicates an association of moral values with an ability to practice positive health behaviors [50-53]. Instead of approaching healthy lifestyles from this neoliberal meritocratic perspective, creating a more health-promoting environment on social media may garner greater community awareness of and involvement in healthy behaviors [37,39,54]. A way to achieve this may be to introduce regulation around social media–delivered advertising campaigns, such as limiting the number of times fast-food advertisements can appear on an individual's newsfeed or using fact-checking systems for health-related posts. Policy reforms regarding social media may also help health promotion to reach a wider range of consumers.

Another key finding of this study was the association of guilt with content from health-focused communities on social media, which is well established in the literature [10,18,55,56]. Young adults are increasingly looking toward health-focused communities for diet and lifestyle guidance, which can have serious consequences related to mental health and body image [10,18,55-58]. They often place greater value on appearance than on health and idealize lean physiques formed through restrictive diet and exercise regimes [55,58,59]. Research indicates that visual comparisons with these body ideals can be detrimental to young adults' self-image and lead to poor mental health [13,55,59-61]. Other studies, including the systematic literature review by Rounsefell et al [13], indicate a link between these comparisons and disordered eating behaviors such as dieting or restricting food and overeating [18]. Participants in this study shared their own feelings of guilt when they were unable to follow the advice of health-focused influencers or achieve their health goals. The coexistence of health-focused content with fast-food advertisements on participants' newsfeeds only exacerbated this condition. Previous findings from our Communicating Health project indicate a moral association with dietary behaviors [53]. In combination with this study, these findings suggest that people may perceive health-focused influencers as the angel on one shoulder and fast-food advertisements as the devil on the other shoulder. Those who follow health-focused pages are shown a message that makes them believe that it is more moral to practice healthy behaviors. When they are unable to follow through with these behaviors, for example, owing to the persuasive impact of fast-food advertisements, this may lead to cognitive dissonance and guilt, as seen in some of our participants.

This study also indicated a gendered response to guilt from viewing health content on social media. Female participants more often discussed the detrimental effects of health content and felt discouraged by upward comparisons, whereas male participants found these to be motivational. These findings contribute to an emerging conversation regarding the impact of health-focused social media content on different genders

XSL•FO
RenderX

[18,62,63]. Women are often perceived to be more vulnerable to the negative impacts of health-focused content on social media, as our own study indicates, and have previously been shown to access diet-related and exercise-related social media posts more commonly than men [64]. However, a growing body of literature suggests that these notions may be caused by gender norms that reduce the likelihood of men openly sharing their experiences with negative body image [18,62,63]. Male participants in the qualitative study by Easton et al [18] revealed negative impacts similar to those experienced by women, a pattern further indicated by a recent cross-sectional survey by DiBisceglie et al [62]. Moreover, a recent study identified that men were featured and objectified in health and fitness content on social media almost as often as women [65]. Further research is needed to clarify whether a meaningful gender-based difference exists in the way that web-based health-focused content is consumed. However, care should be taken to ensure that future social media–delivered health interventions avoid appearance-based health messages to protect young adults' mental health.

Limitations

This study had some limitations. As our data collection was completed during the examination period of Australian Universities, challenges regarding recruitment and participant dropout emerged. Consequently, our participants may not be generally representative of the Australian population. Our sample also included more women than men and a large proportion of students and young adults who were well educated. Our analysis technique included searching for commonalities among the data, which may mean that the experiences of women and students were captured more strongly than those of others. However, we also spent time in contrasting discrepant cases to ensure that less common but still important themes were captured. The conversational design of the forum also may have introduced groupthink, social comparison bias, and recall bias. The dropout rate and different numbers of participants completing each forum may indicate participant fatigue in completing the web-based conversations over an extended period. Moreover, participants may have defined positive or negative health behaviors differently from each other owing to the subjective nature of the topic. Finally, this study was conducted at a particular time with a particular group of Australian young adults. Social media, among other technologies, evolves rapidly. Hence, further research will be necessary as the platforms grow and change.

Conclusions

This study contributed to a greater understanding of the role of social media in health behavior change among young adults. Social factors play a key role in prompting positive health behavior changes. Future studies should develop a greater understanding of social interactions and peer networks in a web-based environment to guide the development of integrated health promotion techniques. The persuasive effect of social media on participants' health behaviors was largely attributable to advertisements and constant exposure to content. This study suggests that young adults view health as an individual responsibility and place great value on self-discipline. A shift toward minimizing external pressures through policy changes and regulation of advertisements needs to be encouraged. Policy reform may also assist health promotion in reaching social media users who are disinterested in health. Finally, future social media–delivered health interventions need to be mindfully developed to ensure that they do not further elicit guilt among social media users.

Acknowledgments

The authors wish to acknowledge the contributions of Associate Professor Catherine Lombard (deceased) and Helen Truby for the planning and conceptualizing the Communicating Health project. The authors would like to acknowledge Karen Klassen and Mike Reid for their work on the web-based conversations. Michelle Raggatt assisted in data analysis. The authors would like to thank all the individuals who participated in the web-based conversations and the market research field house and other members of the research team for their work on the web-based conversations. The Communicating Health project was funded by the Australian National Health and Medical Research Council Targeted Call for Research *Communicating Health: Engaging and retaining young adults in interventions to improve eating behaviours and health outcomes* (GNT1115496). CJCW was funded by the National Health and Medical Research Council Early Career Fellowship. MSCL was funded by the National Health and Medical Research Council Career Development Fellowship.

Authors' Contributions

VF and AM analyzed and interpreted the data for the manuscript drafted by VF. CJCW and MSCL assisted with the interpretation of the findings. TM, LB, and MSCL were involved in the development of the Communicating Health project and design of the web-based conversations. All the authors contributed to the conceptual design and editing of the manuscript.

Conflicts of Interest

None declared.

References

1. Global action plan for the prevention and control of NCDs 2013-2020. World Health Organization. 2013. URL: https://www.who.int/publications/i/item/9789241506236 [accessed 2022-05-09]

XSL•FO
RenderX

2. National health survey: first results, 2017-18 financial year. Australian Bureau of Statistics. 2017. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2022-05-09]
3. Winpenny EM, van Sluijs EM, White M, Klepp KI, Wold B, Lien N. Changes in diet through adolescence and early adulthood: longitudinal trajectories and association with key life transitions. Int J Behav Nutr Phys Act 2018 Sep 10;15(1):86 [FREE Full text] [doi: 10.1186/s12966-018-0719-8] [Medline: 30200990]
4. Arnett JJ. Emerging adulthood. A theory of development from the late teens through the twenties. Am Psychol 2000 May;55(5):469-480. [Medline: 10842426]
5. Deliens T, Clarys P, De Bourdeaudhuij I, Deforche B. Determinants of eating behaviour in university students: a qualitative study using focus group discussions. BMC Public Health 2014 Jan 18;14:53 [FREE Full text] [doi: 10.1186/1471-2458-14-53] [Medline: 24438555]
6. Cha E, Crowe JM, Braxter BJ, Jennings BM. Understanding how overweight and obese emerging adults make lifestyle choices. J Pediatr Nurs 2016;31(6):e325-e332 [FREE Full text] [doi: 10.1016/j.pedn.2016.07.001] [Medline: 27496826]
7. Molenaar A, Choi TS, Brennan L, Reid M, Lim MS, Truby H, et al. Language of health of young Australian adults: a qualitative exploration of perceptions of health, wellbeing and health promotion via online conversations. Nutrients 2020 Mar 25;12(4):887 [FREE Full text] [doi: 10.3390/nu12040887] [Medline: 32218118]
8. Howse E, Hankey C, Allman-Farinelli M, Bauman A, Freeman B. 'Buying salad is a lot more expensive than going to McDonalds': young adults' views about what influences their food choices. Nutrients 2018 Jul 30;10(8):996 [FREE Full text] [doi: 10.3390/nu10080996] [Medline: 30061513]
9. Pelletier JE, Graham DJ, Laska MN. Social norms and dietary behaviors among young adults. Am J Health Behav 2014 Jan;38(1):144-152 [FREE Full text] [doi: 10.5993/AJHB.38.1.15] [Medline: 24034689]
10. Vaterlaus JM, Patten EV, Roche C, Young JA. #Gettinghealthy: the perceived influence of social media on young adult health behaviors. Comput Human Behav 2015 Apr;45:151-157. [doi: 10.1016/j.chb.2014.12.013]
11. Chau MM, Burgermaster M, Mamykina L. The use of social media in nutrition interventions for adolescents and young adults-a systematic review. Int J Med Inform 2018 Dec;120:77-91 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.10.001] [Medline: 30409348]
12. Papasolomou I, Melanthiou Y. Social media: marketing public relations' new best friend. J Promot Manag 2012 Jul;18(3):319-328. [doi: 10.1080/10496491.2012.696458]
13. Rounsefell K, Gibson S, McLean S, Blair M, Molenaar A, Brennan L, et al. Social media, body image and food choices in healthy young adults: a mixed methods systematic review. Nutr Diet 2020 Feb;77(1):19-40 [FREE Full text] [doi: 10.1111/1747-0080.12581] [Medline: 31583837]
14. Yellow social media report 2018. Part one - consumers. Yellow. 2018. URL: https://www.yellow.com.au/wp-content/uploads/2018/06/Yellow-Social-Media-Report-2018-Consumer.pdf [accessed 2022-05-09]
15. Freeman B, Kelly B, Vandevijvere S, Baur L. Young adults: beloved by food and drink marketers and forgotten by public health? Health Promot Int 2016 Dec;31(4):954-961. [doi: 10.1093/heapro/dav081] [Medline: 26276799]
16. Freberg K, Graham K, McGaughey K, Freberg LA. Who are the social media influencers? A study of public perceptions of personality. Public Relat Rev 2011 Mar;37(1):90-92. [doi: 10.1016/j.pubrev.2010.11.001]
17. Uzunoğlu E, Misci Kip S. Brand communication through digital influencers: leveraging blogger engagement. Int J Inf Manag 2014 Oct;34(5):592-602. [doi: 10.1016/j.ijinfomgt.2014.04.007]
18. Easton S, Morton K, Tappy Z, Francis D, Dennison L. Young people's experiences of viewing the fitspiration social media trend: qualitative study. J Med Internet Res 2018 Jun 18;20(6):e219 [FREE Full text] [doi: 10.2196/jmir.9156] [Medline: 29914859]
19. Penders B. Why public dismissal of nutrition science makes sense: post-truth, public accountability and dietary credibility. Br Food J 2018;120(9):1953-1964 [FREE Full text] [doi: 10.1108/BFJ-10-2017-0558] [Medline: 30581197]
20. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What people "Like": analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
21. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
22. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MS. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
23. Barklamb AM, Molenaar A, Brennan L, Evans S, Choong J, Herron E, et al. Learning the language of social media: a comparison of engagement metrics and social media strategies used by food and nutrition-related social media accounts. Nutrients 2020 Sep 16;12(9):2839 [FREE Full text] [doi: 10.3390/nu12092839] [Medline: 32948033]
24. Lombard C, Brennan L, Reid M, Klassen KM, Palermo C, Walker T, et al. Communicating health-optimising young adults' engagement with health messages using social media: study protocol. Nutr Diet 2018 Nov;75(5):509-519. [doi: 10.1111/1747-0080.12448] [Medline: 30009396]

XSL•FO
RenderX

25. Pink S, Horst H, Postill J, Hjorth L, Lewis T, Tacchi J. Digital Ethnography: Principles and Practice. London, UK: Sage Publications; 2016.
26. Brennan L, Fry ML, Previte J. Strengthening social marketing research: harnessing “insight” through ethnography. Australas Mark J 2015;23(4):286-293. [doi: 10.1016/j.ausmj.2015.10.003]
27. Dessart L, Veloutsou C, Morgan-Thomas A. Consumer engagement in online brand communities: a social media perspective. J Prod Brand Manag 2015;24(1):28-42. [doi: 10.1108/jpbm-06-2014-0635]
28. Drinking-related lifestyles: exploring the role of alcohol in Victorians' lives. VicHealth. 2013. URL: https://www.vichealth.vic.gov.au/media-and-resources/publications/drinking-related-lifestyles [accessed 2022-05-09]
29. Code of professional behaviour. The Research Society. 2017. URL: https://researchsociety.com.au/documents/item/2796 [accessed 2022-05-09]
30. In the world of data.. scale matters accuracy matters. Dynata. 2019. URL: https://www.dynata.com/ [accessed 2022-05-09]
31. Australian federal election: it’s getting political!. Pureprofile. 2019. URL: https://www.pureprofile.com/australian-federal-election-its-getting-political/ [accessed 2022-05-09]
32. Our research services. Student Edge. 2017. URL: https://studentedge.org/youth-research-insight-services [accessed 2022-05-09]
33. 3101.0 - Australian Demographic Statistics, Jun 2016. Australian Bureau of Statistics. 2016. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/7d12b0f6763c78caca257061001cc588/2d2860dfa430d432ca2580eb001335bb!OpenDocument [accessed 2022-05-09]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Lewis-Beck MS, Bryman A, Futing Liao T. The SAGE Encyclopedia of Social Science Research Methods. Thousand Oaks, CA, USA: Sage Research Methods; 2004.
36. Patton MQ. Qualitative Research & Evaluation Methods: Integrating Theory and Practice. 4th edition. Thousand Oaks, CA, USA: Sage Publications; 2014.
37. Strong KA, Parks SL, Anderson E, Winett R, Davy BM. Weight gain prevention: identifying theory-based targets for health behavior change in young adults. J Am Diet Assoc 2008 Oct;108(10):1708-1715 [FREE Full text] [doi: 10.1016/j.jada.2008.07.007] [Medline: 18926139]
38. Grunseit AC, Cook AS, Conti J, Gwizd M, Allman-Farinelli M. "Doing a good thing for myself": a qualitative study of young adults' strategies for reducing takeaway food consumption. BMC Public Health 2019 May 07;19(1):525 [FREE Full text] [doi: 10.1186/s12889-019-6731-3] [Medline: 31064366]
39. Stok FM, de Vet E, de Ridder DT, de Wit JB. The potential of peer social norms to shape food intake in adolescents and young adults: a systematic review of effects and moderators. Health Psychol Rev 2016 Sep;10(3):326-340. [doi: 10.1080/17437199.2016.1155161] [Medline: 26878931]
40. Munt AE, Partridge SR, Allman-Farinelli M. The barriers and enablers of healthy eating among young adults: a missing piece of the obesity puzzle: a scoping review. Obes Rev 2017 Jan;18(1):1-17. [doi: 10.1111/obr.12472] [Medline: 27764897]
41. Pilgrim K, Bohnet-Joschko S. Selling health and happiness how influencers communicate on Instagram about dieting and exercise: mixed methods research. BMC Public Health 2019 Aug 06;19(1):1054 [FREE Full text] [doi: 10.1186/s12889-019-7387-8] [Medline: 31387563]
42. Girls make your move: campaign evaluation research. Department of Health, Australian Government. 2019. URL: https://campaigns.health.gov.au/girlsmove/campaign-evaluation [accessed 2022-05-09]
43. Sweeney L. Health Minister announces urgent investigation into taxpayer-funded campaign working with Instagram influencers. ABC News. 2018. URL: https://www.abc.net.au/news/2018-07-20/health-department-investigating-instgram-influencer-campaign/10016712 [accessed 2022-05-09]
44. Holmberg C, E Chaplin JE, Hillman T, Berg C. Adolescents' presentation of food in social media: an explorative study. Appetite 2016 Apr 01;99:121-129. [doi: 10.1016/j.appet.2016.01.009] [Medline: 26792765]
45. Freeman B, Kelly B, Baur L, Chapman K, Chapman S, Gill T, et al. Digital junk: food and beverage marketing on Facebook. Am J Public Health 2014 Dec;104(12):e56-e64. [doi: 10.2105/AJPH.2014.302167] [Medline: 25322294]
46. Montgomery KC, Grier SA, Chester J, Dorfman L. The digital food marketing landscape: challenges for researchers. In: Williams JD, Pasch KE, Collins CA, editors. Advances in Communication Research to Reduce Childhood Obesity. Berlin, Germany: Springer; 2013:221-242.
47. Naylor RW, Lamberton CP, West PM. Beyond the “Like” button: the impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. J Mark 2012 Nov 01;76(6):105-120. [doi: 10.1509/jm.11.0105]
48. Eslami M, Rickman A, Vaccaro K, Aleyasen A, Vuong A, Karahalios K, et al. "I always assumed that I wasn't really that close to [her]": reasoning about invisible algorithms in news feeds. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15; April 18-23, 2015; Seoul, South Korea p. 153-162. [doi: 10.1145/2702123.2702556]
49. Parker L, Brennan L. Social Marketing and Advertising in the Age of Social Media. Cheltenham, UK: Edward Elgar Publishing; 2020.

XSL•FO
RenderX

50. Coffey J. ‘As long as I’m fit and a healthy weight, I don’t feel bad’: exploring body work and health through the concept of ‘affect’. J Sociol 2014 Feb 04;51(3):613-627. [doi: 10.1177/1440783313518249]
51. Coffey J. 3 Youth, health and morality: body work and health assemblages. In: Kelly P, Pike J, editors. Neo-Liberalism and Austerity: The Moral Economies of Young People’s Health and Well-Being. London, UK: Springer; 2017:69-86.
52. Moore SE. Is the healthy body gendered? Toward a feminist critique of the new paradigm of health. Body Soc 2010;16(2):95-118. [doi: 10.1177/1357034x10364765]
53. Brennan L, Klassen K, Weng E, Chin S, Molenaar A, Reid M, et al. A social marketing perspective of young adults' concepts of eating for health: is it a question of morality? Int J Behav Nutr Phys Act 2020 Mar 30;17(1):44 [FREE Full text] [doi: 10.1186/s12966-020-00946-3] [Medline: 32228706]
54. Kuipers G. Cultural narratives and their social supports, or: sociology as a team sport. Br J Sociol 2019 Jun;70(3):708-720 [FREE Full text] [doi: 10.1111/1468-4446.12666] [Medline: 31190396]
55. Raggatt M, Wright CJ, Carrotte E, Jenkinson R, Mulgrew K, Prichard I, et al. "I aspire to look and feel healthy like the posts convey": engagement with fitness inspiration on social media and perceptions of its influence on health and wellbeing. BMC Public Health 2018 Aug 10;18(1):1002 [FREE Full text] [doi: 10.1186/s12889-018-5930-7] [Medline: 30097034]
56. Perloff RM. Social media effects on young women’s body image concerns: theoretical perspectives and an agenda for research. Sex Roles 2014;71(11-12):363-377. [doi: 10.1007/s11199-014-0384-6]
57. Tiggemann M, Zaccardo M. "Exercise to be fit, not skinny": the effect of fitspiration imagery on women's body image. Body Image 2015 Sep;15:61-67. [doi: 10.1016/j.bodyim.2015.06.003] [Medline: 26176993]
58. Jong ST, Drummond MJ. Exploring online fitness culture and young females. Leis Stud 2016 May 8;35(6):758-770. [doi: 10.1080/02614367.2016.1182202]
59. Boepple L, Ata RN, Rum R, Thompson JK. Strong is the new skinny: a content analysis of fitspiration websites. Body Image 2016 Jun;17:132-135. [doi: 10.1016/j.bodyim.2016.03.001] [Medline: 27045871]
60. Want SC. Meta-analytic moderators of experimental exposure to media portrayals of women on female appearance satisfaction: social comparisons as automatic processes. Body Image 2009 Sep;6(4):257-269. [doi: 10.1016/j.bodyim.2009.07.008] [Medline: 19716779]
61. Brannan ME, Petrie TA. Moderators of the body dissatisfaction-eating disorder symptomatology relationship: replication and extension. J Couns Psychol 2008;55(2):263-275. [doi: 10.1037/0022-0167.55.2.263]
62. DiBisceglie S, Arigo D. Perceptions of #fitspiration activity on Instagram: patterns of use, response, and preferences among fitstagrammers and followers. J Health Psychol 2021 Jul;26(8):1233-1242 [FREE Full text] [doi: 10.1177/1359105319871656] [Medline: 31455118]
63. Palmer L. ‘‘Poppin’ bottles, getting wheysted.” Exploring young men’s engagement with fitspiration content and its consequential influences on attitudes and behaviour. J Promot Commun 2015:425-445.
64. Carrotte ER, Vella AM, Lim MS. Predictors of "Liking" three types of health and fitness-related content on social media: a cross-sectional study. J Med Internet Res 2015 Aug 21;17(8):e205 [FREE Full text] [doi: 10.2196/jmir.4803] [Medline: 26297689]
65. Carrotte ER, Prichard I, Lim MS. "Fitspiration" on social media: a content analysis of gendered images. J Med Internet Res 2017 Mar 29;19(3):e95 [FREE Full text] [doi: 10.2196/jmir.6368] [Medline: 28356239]

Edited by JMIRPE Office; submitted 14.03.21; peer-reviewed by J Mayoh, J Liu, A Chung; comments to author 27.04.21; revised version received 20.06.21; accepted 30.12.21; published 18.05.22.

Please cite as:
Friedman VJ, Wright CJC, Molenaar A, McCaffrey T, Brennan L, Lim MSC
The Use of Social Media as a Persuasive Platform to Facilitate Nutrition and Health Behavior Change in Young Adults: Web-Based Conversation Study
J Med Internet Res 2022;24(5):e28063
URL: https://www.jmir.org/2022/5/e28063
doi:10.2196/28063
PMID:35583920

©Vanessa J Friedman, Cassandra J C Wright, Annika Molenaar, Tracy McCaffrey, Linda Brennan, Megan S C Lim. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.05.2022. 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.

XSL•FO
RenderX

Original Paper

Characterization of False or Misleading Fluoride Content on Instagram: Infodemiology Study

Matheus Lotto[1,2], DDS, MSc; Tamires Sá Menezes[1], DDS; Irfhana Zakir Hussain[2,3], BTech; Shu-Feng Tsao[2], BSc, MSc; Zahid Ahmad Butt[2], MBBS, MSc, PhD; Plinio P Morita[2,4,5,6,7], BEng, MSc, PhD; Thiago Cruvinel[1], DDS, MSc, PhD

[1]Department of Pediatric Dentistry, Orthodontics and Public Health, Bauru School of Dentistry, University of São Paulo, Bauru, Brazil

[2]School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

[3]Department of Data Science and Business Systems, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, India

[4]Research Institute for Aging, University of Waterloo, Waterloo, ON, Canada

[5]Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

[6]eHealth Innovation, Techna Institute, University Health Network, Toronto, ON, Canada

[7]Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Thiago Cruvinel, DDS, MSc, PhD
Department of Pediatric Dentistry, Orthodontics and Public Health
Bauru School of Dentistry
University of São Paulo
Alameda Dr Octávio Pinheiro Brisolla, 9-75, Vila Universitária
Bauru, 17012-901
Brazil
Phone: 55 14 3235 8318
Fax: 55 14 3223 4679
Email: thiagocruvinel@fob.usp.br

Abstract

Background: Online false or misleading oral health–related content has been propagated on social media to deceive people against fluoride's economic and health benefits to prevent dental caries.

Objective: The aim of this study was to characterize the false or misleading fluoride-related content on Instagram.

Methods: A total of 3863 posts ranked by users' total interaction and published between August 2016 and August 2021 were retrieved by CrowdTangle, of which 641 were screened to obtain 500 final posts. Subsequently, two independent investigators analyzed posts qualitatively to define their authors' interests, profile characteristics, content type, and sentiment. Latent Dirichlet allocation analysis topic modeling was then applied to find salient terms and topics related to false or misleading content, and their similarity was calculated through an intertopic distance map. Data were evaluated by descriptive analysis, the Mann-Whitney *U* test, the Cramer V test, and multiple logistic regression models.

Results: Most of the posts were categorized as misinformation and political misinformation. The overperforming score was positively associated with older messages (odds ratio [OR]=3.293, *P*<.001) and professional/political misinformation (OR=1.944, *P*=.05). In this context, time from publication, negative/neutral sentiment, author's profile linked to business/dental office/news agency, and social and political interests were related to the increment of performance of messages. Although political misinformation with negative/neutral sentiments was typically published by regular users, misinformation was linked to positive commercial posts. Overall messages focused on improving oral health habits, side effects, dentifrice containing natural ingredients, and fluoride-free products propaganda.

Conclusions: False or misleading fluoride-related content found on Instagram was predominantly produced by regular users motivated by social, psychological, and/or financial interests. However, higher engagement and spreading metrics were associated with political misinformation. Most of the posts were related to the toxicity of fluoridated water and products frequently motivated by financial interests.

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e37519) doi:10.2196/37519

KEYWORDS

eHealth; fluorides; infodemiology; information seeking behavior; internet; misinformation; social media; infoveillance; health outcome; dental caries; health information; dental health

Introduction

The analysis of big data originating from people's production and consumption of online dental information can contribute to recognizing the needs of distinct populations, aiding the planning and implementation of public health actions [1,2]. Within this context emerged the concept of infodemiology, defined as "the science of distribution and determinants of information in an electronic medium with the ultimate aim to inform public health and public policy" [3]. Specifically, internet users have adopted social media to perform queries and express their concerns, doubts, and advice about oral health conditions [4,5]. However, while these behaviors are desirable to provide empowerment and autonomy for individuals toward health education and decision-making [6,7], the content overabundance of social network ecosystems poses a challenge to the public to filter relevant posts, which leads to the consumption of false information and, consequently, the development of damaging health beliefs [8-10]. In this way, previous studies demonstrated that Instagram could be a significant source of health information, including several issues such as COVID-19 and vaccination [11], especially considering the increased popularity of this platform in recent years [12,13].

In this scenario, online false or misleading content propagates the discouragement of the consumption of fluoride-containing water and oral care products concerning their relevance, safety, and harmful consequences [14]. Notably, antifluoridation information is broadly shared on social media, deceiving people against fluoride's economic and health benefits [15]. Moreover, some characteristics of these false or misleading posts, such as the sense of innovation and the negative sentiment charges, favor the diffusion of falsehoods in contrast to trustworthiness [16,17]. In parallel, fluoride refusal is a growing phenomenon observed in dental offices, possibly generated or reinforced by online misinformation [18]. Divergently, there is robust scientific evidence on the beneficial effects of fluoridated water, dentifrices, and mouthwashes to prevent the demineralization and promote the remineralization of dental tissues. Fluoride is considered the most effective measure to reduce the incidence and prevalence of dental caries [19-21], which is the most prevalent oral disease worldwide, affecting the permanent and deciduous teeth of approximately 2.3 billion people and 532 million children, respectively [22].

Thus, the adoption of digital strategies to manage the oral health information disorder on social media is mandatory. Toward this end, the aim of this study was to characterize the false or misleading fluoride-related content on Instagram, regarding authors' and posts' features, interaction and spreading metrics, and the sentiment of posts.

Methods

Study Design

This longitudinal and retrospective infodemiology study analyzed and characterized the false or misleading fluoride-related content of 500 English posts on Instagram. A total of 3863 posts ranked by users' total interaction were retrieved by CrowdTangle, of which 641 were screened for the inclusion criteria. All posts were made available on Instagram between August 2016 and August 2021. Two independent investigators (ML and TSM) analyzed these posts qualitatively to define the authors' interests, profile characteristics, content type, and sentiment of posts. Topic modeling methods were applied to find salient terms and topics related to false or misleading fluoride content. Finally, statistical analysis was performed as described in detail below.

Ethics Considerations

This study did not require institutional review board approval from the Council of Ethics in Human Research of Bauru School of Dentistry because federal regulations do not apply to research using publicly available data that does not involve human subjects. It should be emphasized that the raw data presented in this manuscript have been anonymously disclosed in an open data repository [23].

Search Strategy, Data Collection, and Preprocessing Data Set

CrowdTangle is an online analytics and insights tool owned by Meta Inc that enables the study of several social media metrics such as the number of posts, data, profile information, type of posts, total interaction (sum of the number of likes, comments, and views in a post), and overperforming score through specific keywords. It is also possible to access posts from distinct periods, languages, and social media, besides ranking them into various measures.

The overperforming score is a post's performance regarding its actual interaction divided by its expected interaction according to the number of followers of the author's profile (ie, how many ordinary followers the post reached). In this way, positive scores are associated with good performance posts, reaching a larger user's number than simply the number of the author's followers, and negative scores convey the opposite. Briefly, the algorithm of CrowdTangle generates benchmarks to identify these expected values using the last 100 posts from a given account. For this calculation, the top and bottom 25% posts are dropped and then the mean number of interactions are calculated with the middle 50% of posts in different time intervals (15 minutes old, 60 minutes old, 5 hours old, etc). Subsequently, when the account in question publishes a new post, the platform compares the post metrics to the calculated average and multiplies the difference by the weights in each dashboard [24].

XSL•FO
RenderX

The search strategy ("fluoride free"+"fluoride-free") was defined from exploratory analyses of hashtags and terms related to a higher volume of posts that discouraged fluoride use on Instagram. A data set related to 3863 posts was downloaded as a CSV file on September 15, 2021, regarding specific language (English) and time frame (August 2016 to August 2021), and ranked by total interaction. The period for the collection was determined from the availability of data observed in a preliminary analysis using the search strategy on CrowdTangle and the number of worldwide Instagram users [25]. Furthermore, posts were ranked by total interaction to guarantee the inclusion of those accessed by a considerable volume of Instagram users (ie, those influencing a number of individuals not relativized by the potential of authors to achieve an audience).

Before the qualitative and natural language processing analyses, the raw data set was preprocessed in two ways depending on the type of investigation. First, the data set was screened to obtain a feasible number of posts (n=500), enabling a robust qualitative manual evaluation to feed artificial intelligence–based models, and preventing expected mischaracterization associated with automated tools. Thus, an investigator (ML) read a sample of collected posts (n=641) in full to obtain a list of the first 500 posts ranked by total interaction that satisfied the following inclusion criterion: nonrepeated false or misleading content published in English. The investigator excluded 139 posts due to repetition and 2 posts that were not published in English. It is noteworthy that this process aimed to characterize posts containing false or misleading content with the highest engagement rates on Instagram.

To ensure the quality of topic modeling analysis, another investigator (IZH) performed an additional preprocessing of the words of 500 selected posts, removing symbols, special characters, punctuations, URLs, numbers, personal pronouns, and keywords of the search strategy.

Data Analysis

Qualitative Analysis

The false or misleading fluoride posts were characterized through passive qualitative analysis [26], examining information patterns and interaction metrics. This approach was directed by the most accepted definitions of the categories of information disorder, as follows: (1) misinformation, defined as false information determined based on a grounding of truth and applies only to informationally oriented content [27-29]; (2) fake news, defined as intentionally misleading and biased representational information for the benefit of the messenger sender, which contains false information, with or without a blend of one or more components of omitted important information, a decontextualized content, misleading headlines, or clickbait [30]; (3) disinformation, defined as information that is false and deliberately created to harm a person, social group, organization, or country [27,28]; and (4) conspiracy theories, which are attempts to explain the ultimate causes of significant social and political events and circumstances with claims of secret plots by two or more powerful actors [31].

Additionally, false or misleading online content can be motivated by distinct types of interest such as financial (profiting from information disorder through advertising), political (attempts to influence public opinion due to political positions), social (connecting with a particular group online or offline), and psychological (seeking prestige or reinforcement) [27]. The identification of specific motivations could be a reliable and objective indicator of authors' intentionality, regarding that its determination is only based on the subjective judgment of online content founded on researchers' perspectives [27]. However, according to Poe's law, the clues left by content makers are often inadequate to differentiate between honest and dishonest mistakes (ie, the authors' intentions to deliberately produce or share misleading content to deceive people cannot be categorically identified) [8,32]. Regarding the aforementioned difficulties to establish the specific type of information disorder, misinformation was characterized by two trained and calibrated investigators (ML and TSM) (intraclass correlation coefficient for absolute concordance varying from 0.85 to 0.92), according to the following criteria: author's profile (regular users, business, dental office, or news agency), type of content (commercial or noncommercial), author's interest (social, psychological, financial, and/or political), and sentiment (negative, neutral, or positive). Commercial content was detected when associated with a business, dental office, and news agency, or with regular users identified as influencers for promoting the sales of dental products. Both investigators were trained by the discussion of representative characteristics of posts. The calibration of individual judgment criteria was confirmed by the independent classification of 10% of posts (n=50). The posts that investigators divergently qualified were reassessed until consensus. Additionally, the combination of the author's profile (dental office or others) and the detection of political interests (yes or no) defined the categories of information disorder, grouped as misinformation (posts from regular users without political interests), professional misinformation (posts from a dental office without political interests), and political misinformation (posts from authors with political interests).

Natural Language Processing

Topic modeling is an unsupervised machine learning method that is effectively used to identify patterns within a large corpus of unstructured documents, as previously observed in the health information area [33,34]. Interestingly, researchers who apply unsupervised algorithms do not need to previously define issues in topic modeling, corroborating with the automatized evaluation of social media data sets [35]. Besides a faster analysis, this process allows for identification that would not have been achieved by manual inspection because it is less prone to human biases [36].

We applied latent Dirichlet allocation (LDA) topic modeling using Python 3 in a Google Colab interface to determine the main salient terms and topics from the studied data set, examining the relationship between similar and different content. Synthetically, LDA is a probabilistic and word count–based model that analyzes the frequency of words to determine distinct topics [33]. Given the number of topics *K*, LDA algorithms may generate a keyword list that is most relevant to each topic individually. Although this analysis does not provide a complete

XSL•FO
RenderX

meaning of social media posts, it can contribute to a good overview of issues, facilitating data interpretation [37]. A detailed description of the LDA model is provided elsewhere [38].

We defined the ideal number of topics based on the metric proposed by Nikolenko et al [39] for qualitative studies. Thus, a higher coherence score represents topic modeling with better quality, simplifying the interpretation of outputs. In this way, the coherence values were computed for *K* topics, where *K* ranges from 2 to 50, before eventually narrowing down the consideration range to 3-15 topics. We then carefully examined the models with the highest coherence values and selected that with the most significant score [35,40]. Finally, the topics' distances were calculated to establish their similarity through an intertopic distance map.

Statistical Analysis

Statistical analysis was performed using the Statistical Package for Social Sciences (v. 21.0). First, the variables were dichotomized as follows: time from publication (≤859 or >859 days), categories of information disorder (misinformation or professional/political misinformation), authors' profile (regular users or business/dental office/news agency), sentiment (negative/neutral or positive), type of content (commercial or noncommercial), type of publication (video or photo), total interaction (≤1179 or >1179), overperforming score (≤1.38 or >1.38). The continuous variables were dichotomized from their median values. Additionally, dental offices, news agency, and business profiles were dichotomized on the same side because of their common financial background.

The data normality and homogeneity were determined through the Kolmogorov-Smirnov test and Levene test, respectively. Subsequently, as data were nonnormally distributed, the comparison of total interaction and overperforming score of dichotomized variable groups was performed by the Mann-Whitney *U* test. The differences in the distribution of dichotomized variables according to the categories of information disorder were assessed by the Cramer V test.

Additionally, multiple logistic regression models were developed to evaluate the association of overperforming scores and total interaction with distinct variables. Only factors with significant Wald statistics in the simple analyses were included in the multiple regression models. For all analyses, $P<.05$ was considered significant.

Results

As shown in Table 1, in general, the posts were predominantly commercial, produced by regular users, expressing positive sentiment, and published as an album/photo. The types of interests identified among the 500 selected posts were social (n=500, 100.0%), psychological (n=492, 98.4%), financial (n=421, 84.2%), and political (n=79, 15.8%). Considering the specific interests and authors' profiles, the investigators categorized the posts as misinformation (n=413, 82.6%), political misinformation (n=79, 15.8%), and professional misinformation (n=8, 1.6%).

Table 1 presents the comparison of total interaction and overperforming scores with the distinct dichotomized variable groups. A significantly higher number of total interaction was found for noncommercial content items, whereas a significantly higher overperforming score was detected for >859 days, professional/political misinformation, business/dental office/news agency profiles, and negative/neutral sentiment.

XSL•FO
RenderX

Table 1. Comparison of total interaction and overperforming scores between dichotomized variable groups.

Variable	Posts (N=500), n (%)	Total interaction		Overperforming score		*P* values[a]	
		Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)	Total interaction	Overperforming score
Time from publication						.64	<.001
≤859 days	250 (50.0)	2720 (5031)	1149 (1265)	1.95 (7.48)	1.10 (3.22)		
>859 days	250 (50.0)	2292 (3404)	1222 (1633)	6.15 (18.27)	1.89 (3.57)		
Types of interest						.36	.01
Social, financial, and psychological	421 (84.2)	2468 (4168)	1160 (1473)	4.23 (15.20)	1.33 (4.63)		
Social and political	79 (15.8)	2710 (4449)	1271 (1138)	3.11 (5.15)	1.98 (4.46)		
Author's profile						.46	<.001
Regular users	296 (59.2)	2701 (4759)	1189 (1486)	1.65 (13.79)	1.06 (3.05)		
Business/dental office/news agency	204 (40.8)	2224 (3511)	1155 (1278)	7.54 (13.84)	3.59 (6.40)		
Sentiment						.71	.004
Negative/neutral	77 (15.4)	2699 (5003)	1263 (1118)	3.54 (5.66)	2.30 (4.62)		
Positive	423 (84.6)	2471 (4160)	1164 (1490)	4.15 (15.14)	1.33 (4.52)		
Type of content						.009	.54
Noncommercial	95 (19.0)	3408 (5687)	1554 (2240)	1.98 (4.36)	1.63 (1.63)		
Commercial	405 (81.0)	2295 (3878)	1144 (1279)	4.54 (15.49)	1.35 (5.02)		
Type of publication						.54	.61
Video	45 (9.0)	2835 (4090)	1218 (2017)	2.26 (3.95)	1.70 (4.96)		
Photo	455 (91.0)	2474 (4319)	1174 (1414)	4.23 (14.72)	1.38 (4.77)		

[a]Mann-Whitney *U* test (*P*<.05 considered statistically significant).

Table 2 summarizes the distribution of distinct dichotomized variable groups according to the categories of information disorder. Accordingly, the overperforming score and noncommercial content were significantly higher among professional misinformation and political misinformation groups. Furthermore, political misinformation was frequently posted by regular users with negative/neutral sentiment. By contrast, misinformation commonly presented commercial content with positive feelings.

XSL•FO
RenderX

Table 2. Distribution of dichotomized variable groups according to the categories of information disorder.

Variable	Misinformation (n=413), n (%)	Professional misinformation (n=8), n (%)	Political misinformation (n=79), n (%)	ϕ	*P* value[a]
Time from publication				0.79	.21
≤859 days	209 (50.6)	6 (75.0)	35 (44.3)		
>859 days	204 (49.4)	2 (25.0)	44 (55.7)		
Overperforming				0.149	.004
≤1.38	221 (53.5)	1 (12.5)	30 (37.9)		
>1.38	192 (46.5)	7 (87.5)	49 (62.1)		
Author's profile				0.154	.003
Regular users	247 (59.8)	0 (0)	49 (62.1)		
Business/dental office/news agency	166 (40.2)	8 (100)	30 (37.9)		
Sentiment				0.712	<.001
Negative/neutral	18 (4.3)	0 (0)	59 (74.7)		
Positive	395 (95.7)	8 (100)	20 (25.3)		
Total interaction				0.061	.39
≤1179	212 (51.3)	4 (50.0)	34 (43.1)		
>1179	201 (48.7)	4 (50.0)	45 (56.9)		
Type of content				0.493	<.001
Noncommercial	42 (10.2)	6 (75.0)	47 (59.5)		
Commercial	371 (89.8)	2 (25.0)	32 (40.5)		
Type of publication				0.136	.01
Video	31 (7.5)	0 (0)	14 (17.7)		
Photo	382 (92.5)	8 (100)	65 (82.3)		

[a]Cramer V test (*P*<.05 considered significant).

Table 3 displays the results of the multiple logistic regression model for overperforming score. Overperforming was positively associated with older posts and professional/political misinformation. Notably, total interaction did not show significant Wald statistics for any factor in the simple analysis.

Table 3. Multiple logistic regression model for overperforming score (>1.38).

Variable	B[a] (SE)	Wald statistic	OR[b] (95% CI)	*P* value
Time from publication (>859 days)	1.192 (0.189)	39.84	3.293 (2.274-4.768)	<.001
Information disorder (professional/political misinformation)	0.664 (0.336)	3.900	1.944 (1.005-3.758)	.05
Sentiment (positive)	–0.143 (0.335)	0.163	0.867 (0.434-1.731)	.69
Constant (y-intercept)	–0.605 (0.367)	2.717	0.546	.10

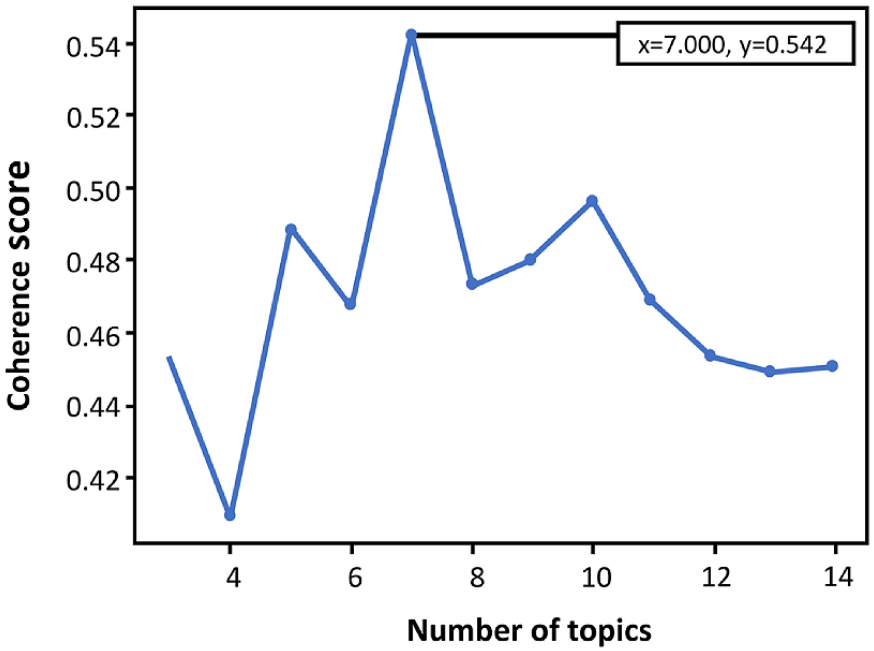
[a]Unstandardized coefficient.

[b]OR: odds ratio.

We adopted an exploratory process to select the topic modeling algorithm with the best performance concerning the coherence score. Figure 1 depicts these values from different number of topics, demonstrating the most significant value for 7 topics (0.54).

XSL•FO
RenderX

Figure 1. Coherence scores for distinct numbers of topics.



Thus, the LDA algorithm was executed with all posts (N=500) through the configuration *K*=7, which generated 7 different fluoride-related topics. Based on the salient keywords of each topic, we attributed a brief description to determine their meaning and subsequently stratified them regarding the main issues, as presented in Table 4. Figure 2 shows the topics' distances to establish their similarity through an intertopic distance map. There was higher proximity of topics 3, 4, and 5; a similarity between topics 1 and 7; and a considerable distance of topics 2 and 6 from the others. Overall, the topics that emerged from the analysis were related to discouraging the consumption of fluoridated products and water by adults and children, justified by their toxicity, using arguments on the improvements of oral health habits (topics 1 and 7), side effects of fluoride (topic 2), the use of dentifrice containing natural and/or vegan ingredients (topics 3, 4, and 5), and propaganda of fluoride-free oral care products (topic 6).

The most representative words of each topic were employed to determine its issues, depending on the specific context of posts, as verified in the manual analysis. For example, the word "giveaway" was linked to posts about dentifrices containing natural and/or vegan ingredients because several authors promote draws of this kind of products, as follows: "It is GIVEAWAY TIME! Baby care is simplified with Dr. Brown's wide range of health and hygiene products."

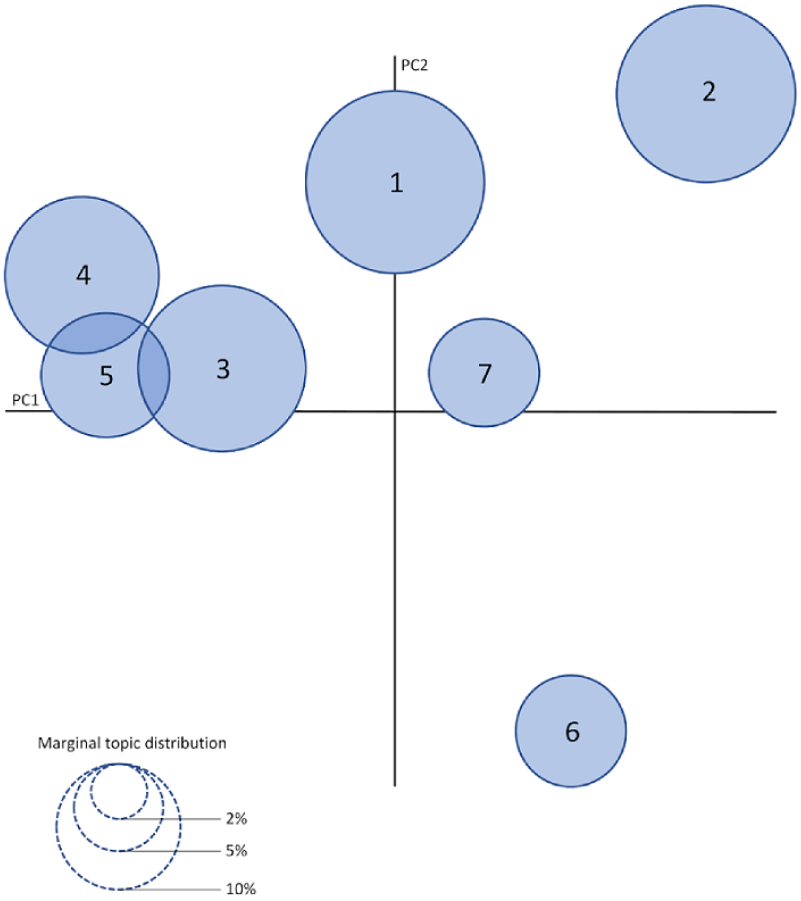
XSL•FO
RenderX

Table 4. Fluoride-related salient topics stratified according to number of posts, most frequent words, issues, and examples.

Topic	Posts, n	Most frequent words	Issues	Examples
1	115	Love, Day, Use, Product, Body, Skin, Time, Natural, Get, Feel, Help, Work, Life, Know, Try	Improvements of oral health habits	"Do your kids enjoy brushing their teeth?! My boys used to fight it until we made it a fun routine!! @grinnatural has now become part of our routine and not only has it helped our kids oral care but has also become a fun activity they look forward to!"
2	62	Water, Drink, Health, Body, Level, Use, Filter, Pineal Gland, Study, High, Know, Brain, Cause, Bone, Source	Side effects of fluoride	"Intentional poisoning of the municipal water sources with toxic fluoride and other toxins/heavy metals of primary source for pineal gland calcification"
3	32	New, Ingredient, Love, Ad, Clean, Formula, Kid, Fresh, Target, Adult, Know, Try, Smile, Toothpaste, Flavor	Use of dentifrice containing natural/vegan ingredients	"Brush-brush the germs away from your baby's teeth with the help of Mee Mee's Fluoride Free Strawberry Flavour Toothpaste"
4	93	Natural, Whiten, Charcoal, Product, Smile, Activate Charcoal, Use, Vegan, Giveaway, Ingredient, White, Toothbrush, Winner, Follow, Coconut	Use of dentifrice containing natural/vegan ingredients	"The Grounded Activated Charcoal Teeth Powder is a 100% natural and fluoride free teeth whitening formula to brighten your teeth shade, remove plaque, cleanse the mouth, remove toxins & make your mouth feel sparkling clean"
5	77	Kid, Brush, Brush Tooth, Love, Baby, Fun, Toothbrush, Fruit, Natural, Ad, Flavor, Routine, Start, Child, Safe Swallow	Use of dentifrice containing natural/vegan ingredients	"#ad Chloe's favorite part of her morning routine is brushing her teeth. Thankfully @toms_of_maine makes brushing her teeth fun with their Silly Strawberry toothpaste. Chloe loves the delicious taste and I love that it's natural free from artificial flavors, colors and preservatives"
6	93	Oral Care, Gum, Mouth, Product, Mouthwash, Bacteria, Disease, Cavity, Oral, Plaque, Bad, Natural, @Garnensgarden, Breath, Garnens Garden	Propaganda of fluoride-free oral care products	"Make the Switch, to an all-natural oral care products from Garners Garden (@garnersgarden)! Protect your gums and teeth from cavities and bad bacteria!"
7	28	Organic, Use, Tongue, Add, Oil, Healthy, Daily, Toxin, Routine, Clean, Coconut Oil, Brush, Day, Tap, and Antibacterial	Improvements of oral health habits	"How many of you guys Oil Pull? It's one of my favorite ways to detox and keep my teeth healthy/white"

XSL•FO
RenderX

Figure 2. Intertopic distance map of the topic modeling analysis. Note that the bubbles are denominated according to the number of the specific topic. PC: principal component.



Discussion

Principal Findings and Comparison With Prior Work

These findings indicate that the predominant false or misleading fluoride Instagram posts were categorized as misinformation (n=413) and political misinformation (n=79). In this context, several characteristics were related to the increment of overperforming scores of messages, such as time from publication, negative or neutral sentiment, business/dental office/news agency author's profile, and social and political interests. In particular, older messages (odds ratio [OR]=3.29) and professional/political misinformation (OR=1.94) were associated with better performance of spreading among Instagram users. Remarkably, commercial content was significantly more prevalent in the misinformation category than in the professional and political misinformation categories. Furthermore, regular users preponderantly published political misinformation presenting negative or neutral sentiment, whereas misinformation was linked to positive commercial posts. The messages generally addressed the toxicity of fluoridated products and water, focusing on improving oral health habits, side effects of fluoride, dentifrice containing natural and/or vegan compounds, and propaganda of fluoride-free oral care products. Although previous studies have analyzed fluoride-related information on social media, including Instagram [13-15,17,41], this study differs regarding only focusing on analyses of false or misleading fluoride information, identified based on contemporary concepts and methods on information disorder.

From these outcomes, we confirmed that oral health information seekers engage more with political fluoride misinformation, even after excluding the influence of time as a confounding factor. Indeed, social media consumers tend to connect with others similar to themselves regarding political ideology [42]. People motivated by specific political overviews, influenced by personal characteristics such as beliefs and values, are predisposed to be more interactive with congruent arguments and assimilate them uncritically (confirmation bias) [43,44]. Thus, greater political homophily is associated with increased user interaction since it reinforces similar ideologies [45]. It is important to note that individuals are susceptible to believing and sharing misinformation regardless of their underlying political creed [44].

Moreover, LDA topic modeling categorized most of the political misinformation in topic 2, covering the possible side effects from fluoride toxicity, as exemplified by the following posts:

> *over three hundred studies have found that fluoride is literally a neurotoxin*
>
> *fluoridated water provides no benefits, only risks. Babies given fluoridated water in their formula may have reduced IQ scores.*

XSL•FO
RenderX

This demonstrates that Instagram users were strongly influenced by concerns and fears surrounding fluoridated products and the water supply, interacting with negative sentiment posts that emphasized the adverse health aspects of fluoride. These outcomes are in agreement with posts of Twitter users [17].

The positive impact of the time of availability of posts on overperforming scores is an expected result because users have more opportunities to access these posts in comparison to more recent posts. Likewise, authors' profiles linked to economic activities, such as companies, dental offices, and news media, usually structure their messages to attract customers, besides probably paying money to promote their content on Instagram, which increases people's engagement and thus raises content diffusion. Surprisingly, we detected financial interest in most posts, including a substantial portion of regular users (digital influencers) that publicized fluoride-free products. Moreover, several salient topics that emerged from modeling were closely connected to brands. Indeed, the distribution of information disorder often has a close relationship with economic gains [27]. Specifically, our findings suggest that the antifluoridation proposals strongly connect with financial concerns beyond the above-discussed ideological aspects. In this sense, distinct oral care companies have been focused on developing products that meet the individual wishes of consumers, even with the absence of scientific evidence [46].

Practical Implications

These findings can support the development of methods and models to automatically identify false or misleading content items and assess their propagation on social media. In addition, outcomes such as topic modeling can subside the elaboration of eHealth and mobile health fluoride-related educational approaches to guide social media users toward the consumption of adequate online oral health information [47]. In this context, dental professional teams need to be conscious of fluoride-related misinformation toward improving the quality of their relationship with patients. Additionally, universal access to oral health, improving eHealth and electronic literacy, and offering high-quality dental information are desirable to prevent the consumption of deceptive messages. Certainly, policymakers should recognize the negative influence of these false posts on communities, creating guidelines and laws to control the spread of information disorder. Specifically, social media managers should be encouraged to develop mechanisms for screening posts to detect false or misleading content before considering messages eligible for sponsorship, avoiding the dissemination of misinformation. Despite the difficulties in determining the authors' intentions, society needs to start discussing education measures and possible penalties for misinformation propagators, within the confines of democratic values, mainly when disseminated by health professionals.

Limitations

This study has some limitations. First, we collected the sample from a specific search strategy composed of two keywords, limiting the findings' generalization to all false or misleading fluoride content. However, we performed an exploratory analysis to determine the most representative keywords with the greatest spread for the thematic analysis in data collection. Second, the two independent investigators analyzed only 500 posts due to work restrictions associated with human analysis, in accordance with previous dental studies [4]. In addition, the manual labeling of data sets is imperative to training artificial intelligence models for natural processing language tasks, ensuring high accuracy and data generalizability [48]. Third, as previously described, we cannot differentiate misinformation from other types of information disorder because of the incapacity of determining authors' intentionality objectively and precisely [49]. Notwithstanding, the characterization of misinformation was improved, verifying the association of specific interests and authorship with interaction metrics. Fourth, these interpretations were based on content published in English. Although English is the most spoken language worldwide, cultural aspects likely influenced the detection of falsehoods.

Conclusions

False or misleading fluoride posts available on Instagram were predominantly characterized as misinformation produced by regular users motivated by social, psychological, and/or financial interests; however, misinformation with social and political interests was associated with higher engagement and spreading metrics. In general, the content of posts was related to the toxicity of fluoridated water and products, frequently motivated by financial interests.

Acknowledgments

This work was supported by the São Paulo Research Foundation (grants 2019/27242-0 and 2021/03226-6). The authors are grateful to Meta Inc for granting use of the CrowdTangle platform.

Conflicts of Interest

None declared.

References

1. Cruvinel T, Ayala Aguirre PE, Lotto M, Marchini Oliveira T, Rios D, Pereira Cruvinel AF. Digital behavior surveillance: monitoring dental caries and toothache interests of Google users from developing countries. Oral Dis 2019 Jan;25(1):339-347. [doi: 10.1111/odi.12986] [Medline: 30270556]

XSL•FO
RenderX

2. Rizzato VL, Lotto M, Lourenço Neto N, Oliveira TM, Cruvinel T. Digital surveillance: the interests in toothache-related information after the outbreak of COVID-19. Oral Dis 2021 Aug 27:1-10 [FREE Full text] [doi: 10.1111/odi.14012] [Medline: 34448289]
3. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
4. Heaivilin N, Gerbert B, Page JE, Gibbs JL. Public health surveillance of dental pain via Twitter. J Dent Res 2011 Sep;90(9):1047-1051 [FREE Full text] [doi: 10.1177/0022034511415273] [Medline: 21768306]
5. Graf I, Gerwing H, Hoefer K, Ehlebracht D, Christ H, Braumann B. Social media and orthodontics: a mixed-methods analysis of orthodontic-related posts on Twitter and Instagram. Am J Orthod Dentofacial Orthop 2020 Aug;158(2):221-228. [doi: 10.1016/j.ajodo.2019.08.012] [Medline: 32576427]
6. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
7. Barber SK, Lam Y, Hodge TM, Pavitt S. Is social media the way to empower patients to share their experiences of dental care? J Am Dent Assoc 2018 Jun;149(6):451-459. [doi: 10.1016/j.adaj.2018.01.007] [Medline: 29656806]
8. Giglietto F, Iannelli L, Valeriani A, Rossi L. 'Fake news' is the invention of a liar: how false information circulates within the hybrid news system. Curr Sociol 2019 Apr 08;67(4):625-642. [doi: 10.1177/0011392119837536]
9. Strieder AP, Aguirre PEA, Lotto M, Cruvinel AFP, Cruvinel T. Digital behavior surveillance for monitoring the interests of Google users in amber necklace in different countries. Int J Paediatr Dent 2019 Sep;29(5):603-614. [doi: 10.1111/ipd.12500] [Medline: 30920686]
10. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
11. Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. J Med Internet Res 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]
12. Massey PM, Kearney MD, Hauer MK, Selvan P, Koku E, Leader AE. Dimensions of misinformation about the HPV vaccine on Instagram: content and network analysis of social media characteristics. J Med Internet Res 2020 Dec 03;22(12):e21451 [FREE Full text] [doi: 10.2196/21451] [Medline: 33270038]
13. Niknam F, Samadbeik M, Fatehi F, Shirdel M, Rezazadeh M, Bastani P. COVID-19 on Instagram: A content analysis of selected accounts. Health Policy Technol 2021 Mar;10(1):165-173 [FREE Full text] [doi: 10.1016/j.hlpt.2020.10.016] [Medline: 33723502]
14. Mackert M, Bouchacourt L, Lazard A, Wilcox GB, Kemp D, Kahlor LA, et al. Social media conversations about community water fluoridation: formative research to guide health communication. J Public Health Dent 2021 Jun;81(2):162-166. [doi: 10.1111/jphd.12404] [Medline: 33058200]
15. Mertz A, Allukian M. Community water fluoridation on the Internet and social media. J Mass Dent Soc 2014;63(2):32-36. [Medline: 25230407]
16. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
17. Oh HJ, Kim CH, Jeon JG. Public sense of water fluoridation as reflected on Twitter 2009-2017. J Dent Res 2020 Jan;99(1):11-17. [doi: 10.1177/0022034519885610] [Medline: 31682777]
18. Eliacik BK. Topical fluoride applications related posts analysis on Twitter using natural language processing. Oral Health Prev Dent 2021 Jan 07;19(1):457-464. [doi: 10.3290/j.ohpd.b2048359] [Medline: 34546013]
19. Walsh T, Worthington HV, Glenny A, Marinho VC, Jeroncic A. Fluoride toothpastes of different concentrations for preventing dental caries. Cochrane Database Syst Rev 2019 Mar 04;3:CD007868 [FREE Full text] [doi: 10.1002/14651858.CD007868.pub3] [Medline: 30829399]
20. Belotti L, Frazão P. Effectiveness of water fluoridation in an upper-middle-income country: a systematic review and meta-analysis. Int J Paediatr Dent. Preprint posted online on September 26, 2021. [doi: 10.1111/ipd.12928] [Medline: 34564916]
21. Whelton HP, Spencer AJ, Do LG, Rugg-Gunn AJ. Fluoride revolution and dental caries: evolution of policies for global use. J Dent Res 2019 Jul;98(8):837-846. [doi: 10.1177/0022034519843495] [Medline: 31282846]
22. GBD 2017 Oral Disorders Collaborators, Bernabe E, Marcenes W, Hernandez CR, Bailey J, Abreu LG, et al. Global, regional, and national levels and trends in burden of oral conditions from 1990 to 2017: a systematic analysis for the Global Burden of Disease 2017 Study. J Dent Res 2020 Apr;99(4):362-373 [FREE Full text] [doi: 10.1177/0022034520908533] [Medline: 32122215]
23. Lotto M, Menezes TS, Hussain IZ, Tsao SF, Butt ZA, Morita PP, et al. Raw data of the manuscript: Characterization of false or misleading fluoride content on Instagram: Infodemiology study. Figshare. 2022 Jan 31. URL: https://figshare.com/articles/dataset/Raw_data_of_the_manuscript_Characterization_of_misleading_fluoride_information_on_Instagram_/19099733 [accessed 2022-02-22]

XSL•FO
RenderX

24. How is overperforming calculated? CrowdTangle. 2022. URL: https://help.crowdtangle.com/en/articles/1141056-how-is-overperforming-calculated [accessed 2022-02-22]
25. Number of monthly active Instagram users 2013-2021. Statista. 2022. URL: https://www.statista.com/statistics/253577/number-of-monthly-active-instagram-users/#:~:text=Social%20media%20usage%20worldwide,with%20114.9%20million%20active%20users [accessed 2022-02-22]
26. Franz D, Marsh HE, Chen JI, Teo AR. Using Facebook for qualitative research: a brief primer. J Med Internet Res 2019 Aug 13;21(8):e13544 [FREE Full text] [doi: 10.2196/13544] [Medline: 31411143]
27. Wardle C, Derakhshan H. Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. 2017. URL: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html [accessed 2022-02-22]
28. Journalism, 'Fake News' and Disinformation: A Handbook for Journalism Education and Training. UNESCO. 2018. URL: https://en.unesco.org/fightfakenews
29. Molina MD, Sundar SS, Le T, Lee D. “Fake News” is not simply false information: a concept explication and taxonomy of online content. Am Behav Sci 2019 Oct 14;65(2):180-212. [doi: 10.1177/0002764219878224]
30. Lim S. Academic library guides for tackling fake news: a content analysis. J Acad Librariansh 2020 Sep;46(5):102195. [doi: 10.1016/j.acalib.2020.102195]
31. Douglas KM, Uscinski JE, Sutton RM, Cichocka A, Nefes T, Ang CS, et al. Understanding conspiracy theories. Polit Psychol 2019 Mar 20;40(S1):3-35. [doi: 10.1111/pops.12568]
32. Aikin SF. Poe's Law, group polarization, and argumentative failure in religious and political discourse. Soc Semiot 2013 Jun;23(3):301-317. [doi: 10.1080/10350330.2012.719728]
33. Asmussen CB, Møller C. Smart literature review: a practical topic modelling approach to exploratory literature review. J Big Data 2019 Oct 19;6(1):93. [doi: 10.1186/s40537-019-0255-7]
34. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using latent Dirichlet allocation for topic modeling on Twitter. PLoS One 2020;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
35. Pang PC, McKay D, Chang S, Chen Q, Zhang X, Cui L. Privacy concerns of the Australian My Health Record: implications for other large-scale opt-out personal health records. Inf Process Manag 2020 Nov;57(6):102364. [doi: 10.1016/j.ipm.2020.102364]
36. Hagen L. Content analysis of e-petitions with topic modeling: how to train and evaluate LDA models? Inf Process Manag 2018 Nov;54(6):1292-1307. [doi: 10.1016/j.ipm.2018.05.006]
37. Jockers ML, Mimno D. Significant themes in 19th-century literature. Poetics 2013 Dec;41(6):750-769. [doi: 10.1016/j.poetic.2013.08.005]
38. Blei DM. Probabilistic topic models. Commun ACM 2012 Apr 01;55(4):77-84. [doi: 10.1145/2133806.2133826]
39. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. J Inf Sci 2016 Jul 10;43(1):88-102. [doi: 10.1177/0165551515617393]
40. Röder M, Both A, Hinneburg A. Exploring the space of topic coherence measures. 2015 Presented at: 8th ACM International Conference on Web Search and Data Mining; 2015; Shanghai, China p. 339-408. [doi: 10.1145/2684822.2685324]
41. Basch CH, Milano N, Hillyer GC. An assessment of fluoride related posts on Instagram. Health Promot Perspect 2019;9(1):85-88 [FREE Full text] [doi: 10.15171/hpp.2019.11] [Medline: 30788272]
42. Mitchell A, Gottfried J, Kiley J, Matsa KE. Pew Research Center. 2014 Oct 21. URL: https://www.pewresearch.org/journalism/2014/10/21/political-polarization-media-habits/ [accessed 2022-02-22]
43. Kim A, Moravec PL, Dennis AR. Combating fake news on social media with source ratings: the effects of user and expert reputation ratings. J Manag Inf Syst 2019 Aug 04;36(3):931-968. [doi: 10.1080/07421222.2019.1628921]
44. Scherer LD, Pennycook G. Who is susceptible to online health misinformation? Am J Public Health 2020 Oct;110(S3):S276-S277. [doi: 10.2105/AJPH.2020.305908] [Medline: 33001736]
45. Boutyline A, Willer R. The social structure of political echo chambers: variation in ideological homophily in online networks. Polit Psychol 2016 May 05;38(3):551-569. [doi: 10.1111/pops.12337]
46. Bauler LD, Santos CSD, Lima GS, Moraes RR. Charcoal-based dentifrices and powders: analyses of product labels, Instagram engagement, and altmetrics. Braz Dent J 2021;32(2):80-89 [FREE Full text] [doi: 10.1590/0103-6440202104233] [Medline: 34614064]
47. 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] [Medline: 32827598]
48. 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] [Medline: 33724919]

XSL•FO
RenderX

49. Son GHW, Rashid EIA. Classification of information disorder. Khazanah Research Institute. 2022. URL: http://www.krinstitute.org/assets/contentMS/img/template/editor/DP%20-%20Classification%20of%20Information%20Disorder.pdf [accessed 2022-02-22]

Abbreviations

LDA: latent Dirichlet allocation
OR: odds ratio

Edited by A Mavragani; submitted 23.02.22; peer-reviewed by R Ratto Moraes, PCI Pang, L Bouchacourt, J Chen; comments to author 21.03.22; revised version received 01.04.22; accepted 14.04.22; published 19.05.22.

Please cite as:
Lotto M, Sá Menezes T, Zakir Hussain I, Tsao SF, Ahmad Butt Z, P Morita P, Cruvinel T
Characterization of False or Misleading Fluoride Content on Instagram: Infodemiology Study
J Med Internet Res 2022;24(5):e37519
URL: https://www.jmir.org/2022/5/e37519
doi:10.2196/37519
PMID:35588055

©Matheus Lotto, Tamires Sá Menezes, Irfhana Zakir Hussain, Shu-Feng Tsao, Zahid Ahmad Butt, Plinio P Morita, Thiago Cruvinel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2022. 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.

XSL•FO
RenderX

Original Paper

Patient Activeness During Online Medical Consultation in China: Multilevel Analysis

Bolin Cao[1], PhD; Wensen Huang[1], PhD; Naipeng Chao[1], PhD; Guang Yang[1], PhD; Ningzheng Luo[2], EMBA

[1]School of Media and Communication, Shenzhen University, Shenzhen, China

[2]Health 160, Shenzhen, China

Corresponding Author:
Wensen Huang, PhD
School of Media and Communication
Shenzhen University
L7 Building, South Campus, Number 3688, Nanhai Avenue
Nanshan Area
Shenzhen, 518000
China
Phone: 86 17717428439
Email: wshuang19@szu.edu.cn

Abstract

Background: Online medical consultation is an important complementary approach to offline health care services. It not only increases patients' accessibility to medical care, but also encourages patients to actively participate in consultation, which can result in higher shared decision making, patient satisfaction, and treatment adherence.

Objective: This study aims to explore multilevel factors that influence patient activeness in online medical consultations.

Methods: A data set comprising 40,505 patients from 300 physicians in 10 specialties was included for multilevel analysis. Patient activeness score (PAS) was calculated based on the frequency and the proportion of patient discourses to the total frequency of doctor-patient interactions. Intraclass correlation coefficients were calculated to identify between-group variations, and the final multilevel regression model included patient- and physician-level factors.

Results: Patients were not equally active in online medical consultations, with PASs varying from 0 to 125.73. Patient characteristics, consultation behavioral attributes, and physician professional characteristics constitute 3 dimensions that are associated with patient activeness. Specifically, young and female patients participated more actively. Patients' waiting times online (β=–.17; P<.001) for physician responses were negatively correlated with activeness, whereas patients' initiation of conversation (β=.83; P<.001) and patient consultation cost (β=.52; P<.001) in online medical consultation were positively correlated. Physicians' online consultation volumes (β=–.10; P=.01) were negatively associated with patient activeness, whereas physician online consultation fee (β=.03; P=.01) was positively associated. The interaction effects between patient- and physician-level factors were also identified.

Conclusions: Patient activeness in online medical consultation requires more scholarly attention. Patient activeness is likely to be enhanced by reducing patients' waiting times and encouraging patients' initiation of conversation in online medical consultation. The findings have practical implications for patient-centered care and the improvement of online medical consultation services.

(J Med Internet Res 2022;24(5):e35557) doi:10.2196/35557

KEYWORDS

patient; physician; online medical consultation; patient activeness

Introduction

Online medical consultation is increasingly being chosen by patients as an alternative to traditional health care services. It is an electronic format for doctor-patient interaction, connecting both parties of medical services through text, pictures, and videos [1-3]. The development of online medical consultation is a result of rapid technological advancement and broad market demand [2,4,5]. Specifically, online medical consultation has been emphasized during the COVID-19 pandemic to prevent cross-infection or to implement social distancing rules as encouraged by many governments [6]. Given its advantages in

XSL•FO
RenderX

removing temporal and spatial barriers and documenting the medication process, online medical consultation is sometimes more appealing to patients than offline medical encounters [2,7]. According to the global market estimate reports, the market share of online medical consultation was US $3.9 billion in 2020 and is estimated to reach US $16.0 billion by 2026 [8].

Although online medical consultation has been widely recognized for its potential, its strengths remain underappreciated. An overlooked benefit is its encouragement of patient activeness. Patient activeness is an emerging term to describe patients' active participation in managing their health and wellness. Patients often participate in their medical visits by seeking and providing information, asserting their preferences or opinions, and expressing their concerns [9]. Through these formats, patients can gain more knowledge and control over their health, leading to improved health outcomes [7]. Patient activeness has also been found to have significant impacts on shared decision making, patient satisfaction, and treatment adherence [10-12].

Following the patient-centered approach, online medical consultation may empower patients with more opportunities to actively participate in the health care–seeking process. First, patients can express themselves extensively without feeling rushed in online medical consultation. Patients can deliver unlimited messages in the text format without being interrupted by the physician or another patient [1,13]. Second, patients who feel tense and shy in the offline setting might be freer to disclose themselves, given that the absence of a visual audience may allow them to let go of their sense of being watched or embarrassed [13]. Third, patients can access the internet or other resources for additional information in online medical consultation, enhancing their ability to participate in online medical conversations with their physicians [2,7]. The internet can help patients bridge the information gap with their physicians and ask more meaningful questions. As a result of these features of online medical consultation, the online doctor-patient communication process can be transformed from paternalism to partnership, encouraging patient activeness [14].

Traditionally, health providers play a paternalistic role in China, only providing information and treatments they consider necessary and useful for patient recovery [15]. This leads to patients being viewed as passive, rarely offering an opinion, and not participating in medical decision making [16]. Most offline consultations last for less than 5 minutes, and the patient participates only minimally during the consultation [17,18]. Like many other countries, China has seen an increase in the demand for online medical care. According to a report published by consultancy iiMedia Research [4], the market for online medical care in China was US $8.41 billion in 2019. Chinese internet users who use online medical services have increased threefold from 2015 to 2020, with a total of 661 million users. Most online medical consultation platforms offer text consultation, telephone consultation, and video consultation. Telephone consultations and video consultations are both synchronous, while text consultations are asynchronous. However, the most common form of online consultation among patients is text consultation due to several reasons, including the fact that most physicians conduct consultations during their spare time, scheduling appointments for a telephone or video consultation is challenging, and the cost of telephone or video consultations are high [18].

This study aims to investigate factors that influence patient activeness during online medical consultation in China. Despite its popularity, online medical consultation remains in its infancy in China [12]. Patients' active participation in online medical consultation could increase their preference and satisfaction with the service. As online medical consultation involves interpersonal and environmental dynamics [19,20], this study examines the multilevel factors that influence patient participation in online consultation.

Methods

Data Collection and Research Design

Data for this study were collected from a third-party online medical consultation platform, Health 160, a pioneering online medical service provider. Established in 2005, it has served over 170 million patients associated with 610,000 physicians. Anonymous and deidentified consultation data were obtained from this platform, complying with the users' consent to privacy policies. A multistage sampling scheme was conducted, where the top 10 disease specialties were selected on the basis of their popularity first, and then 30 physicians were selected at random within each disease specialty. Between May 2019 and May 2020, these 300 physicians' consultation records were retrieved, yielding 57,378 consultations from 40,505 patients. Given that some patients had more than 1 order for consultation in the sample, the first record for a specific patient was utilized.

A hierarchical data structure was used in this study, with patients nested within physicians within disease specialties. The 40,505 consultation records were considered as the first layer of the research, which included patient characteristics and consultation behavior data, and 300 physician-level data as the second layer of the research.

Ethics Approval

The study protocol was approved by the institutional review board at Shenzhen University (approval number 2020028).

Measure

Patient activeness was measured by the patient activeness score (PAS), which expresses the proportion of patient discourses to the total frequency of doctor-patient communication in online medical consultation. The formula of calculating PAS is as follows:

$$PAS = n \times [n/(n+m)]$$

where n and m refer to the numbers of patient discourses and physician expressions for each consultation, respectively. Given the highly skewed distribution of PAS from 0 to 125.73, a logarithmical transformation was used for each score: $f(x) = \ln(x+0.01)$, which reduced the variability of scores to [–4.61 to 4.83], denoted as logPAS.

First-Level Variables

Patient demographics and the consultation behavioral characteristics constitute the first-level factors of analysis.

XSL•FO
RenderX

Patient demographics include gender and age, and the consultation behavioral characteristics include patients' waiting time for response, patients' initiation of consultation, and patient cost for consultation service. Patients' waiting time for response refers to how long they waited for the doctor to respond after a consultation has been initiated. The time was primarily measured in seconds and then converted to log10: f(x) = log10(x). Patients' initiation of consultation was dummy coded as whether the conversation was started by the patient or the physician (patient = yes, physician = no). In addition, the online consultation services on the Health 160 platform are generally paid, but some free consultation services are provided to promote service use, facilitate user trials, and aid in the prevention and control of the COVID-19 epidemic during the time the data were collected. Thus, the patient cost for consultation service was also dummy coded as paid or free.

Second-Level Variables

In our study design, physician-level factors were considered second-level variables to investigate the effect of group heterogeneity on patient activeness. These factors included physician demographics and professional characteristics. The demographic variables included physician age (in years) and gender (1 = female, 0 = male). Their professional characteristics were indicated by their online consultation volumes and fees. Online consultation volume refers to the number of patients who have consulted the physician online. Given that the online consultation volumes ranged from 3 to 3916, they were log-transformed to values with a base of 10 and the range was 0.48-3.59. The consultation fee was recoded as an ordinal variable, ranging from 0 (0 CNY) to 6 (over 50 CNY [US $7.72]).

Third-Level Variables

Specialty-related factors were deemed as the third-level variable. The 10 different specificities were coded as categorical variables for analysis. The descriptive statistics of first-, second-, and third-level variables are shown in Table 1.

Statistical Analysis

Multilevel regression analysis was used to analyze the nested study design. Generally, multilevel regression analysis is considered convenient for modeling the possible contributions of contextual factors at higher levels. An initial analysis comprising a 3-level model that incorporated patient (Level 1), physician (Level 2), and disease specialty (Level 3) characteristics was first conducted. To identify whether the higher levels were critical to explaining the data, the intraclass correlation coefficient (ICC) was estimated (denoted as ρ), an indication of how much variation in the outcome variable can be explained by between-group variation. The ICC number ranges from 0 to 1, with higher values indicating a greater variance between groups. Statisticians suggested that a variance of 0.059 might be used as an experience criterion to determine whether the between-group difference was large enough to be regarded as above average [21]. At the specialty level, the ICC was evaluated to be 0.018, indicating a low level of clustering effect within disease specialties in which physicians were similar. About 13% of the variation was attributed to variation at the physician level (ie, ICC ρ = [0.017 + 0.105]/[0.017 + 0.105 + 0.848] = 0.126). Thus, a 2-level analysis was applied, ignoring the specialty level, with patients nested within physicians.

In our specific analysis, the effect of each predictor on the outcome variable was first analyzed using univariate regression. Second, 2-level predictors were included successively in the nested models, testing the relative contribution of each with multiple regression analysis. For each independent variable, the variance inflation factors were found to be below 1.80, indicating that the collinearity between independent variables could safely be ignored without experiencing multicollinearity problems.

In addition, several cross-level interactions were identified between physician- and patient-level factors that might influence patient activeness. For instance, patients' waiting time for response was supposed to be related to physician online consultation volumes, patients' initiation of consultation was supposed to be related to physician gender, and patient cost for consultation service was supposed to be related to physician online consultation fee. During the interaction analysis, the centering of the explanatory variables is advantageous when a multilevel model contains interactions, given that it provides a clear interpretation of interaction terms and facilitates computation and convergence [22]. In this study, a grand mean centering, subtracting the mean from all values, was performed for each variable involved in the interactions. All statistical analyses were performed using R, version 3.6.

Results

Sample Characteristics

Of the 40,505 patients included in this study, there are twice as many female patients (n=28,057, 69.27%) as male patients (n=12,448, 30.73%). The median (IQR) proportion score of patient discourse during consultation [$n/(n+m)$] was 0.54 (0.50-0.66). The frequency of patient discourse (n) during consultation ranged from 0 to 178, and the frequency of patient and physician discourses ($n+m$) in total during consultation ranged from 1 to 227.

Among the 300 physicians, there existed slightly more females (n=163, 54%) than males (n=137, 46%). One-third of physicians were attending physicians (97/300, 32.33%), and nearly two-thirds held the higher management titles of chief physicians (81/300, 27.00%) and deputy chief physicians (104/300, 34.67%). The majority of consultations (34,024/40,505, 84%) were initiated by physicians. The median (IQR) patients' waiting time for response was 7350 (1486-23,974) seconds, which is roughly 2 hours. Approximately one-fourth (9820/40,505, 24.4%) of patients received free online consultation. Of the 300 physicians, 19.00% (n=57) did not charge at all and the majority (n=213) did not charge over 50 CNY (US $7.85), whereas 10% (n=30) charged over 50 CNY (US $7.85). As of May 2020, the median (IQR) level of physicians' online consultation volume reached 94.5 (43-211.75) patients (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of first-, second-, and third-level variables.

Characteristics	Values
Patient-level characteristics (N=40,505)	
Patient demographics	
Age, median (IQR)	27 (11-33)
Gender, n (%)	
Male	12,448 (30.73)
Female	28,057 (69.27)
Patient consultation behavioral characteristics	
Patients' waiting time for response (seconds), median (IQR)	7350 (1486-23,974)
Patients' initiation of consultation, n (%)	
Yes	6318 (15.60)
No	34,187 (84.40)
Patient cost for consultation, n (%)	
Paid	30,685 (75.76)
Free	9820 (24.24)
Physician-level characteristics (N=300)	
Physician demographics	
Age, median (IQR)	44 (37-51)
Gender, n (%)	
Male	137 (45.67)
Female	163 (54.33)
Professional title, n (%)	
Chief physician	81 (27.00)
Deputy chief physician	104 (34.67)
Attending physician	97 (32.33)
Other	18 (6.00)
Physician professional characteristics	
Physician online consultation volume, median (IQR)	94.5 (43-211.75)
Physician online consultation fee (CNY), n (%)[a]	
0	57 (19.00)
1-10	12 (4.00)
11-20	94 (31.33)
21-30	52 (17.33)
31-40	12 (4.00)
41-50	43 (14.33)
≥51	30 (10.00)
Disease specialties (N=40,505), n (%)	
Dermatology	7975 (19.69)
Gynecology	6840 (16.89)
Pediatrics	5017 (12.39)
Endocrinology	3982 (9.83)
Traditional Chinese Medicine	3942 (9.73)
Obstetrics	3263 (8.06)

XSL•FO
RenderX

Characteristics	Values
Urology	2918 (7.20)
Stomatology	2562 (6.33)
Psychiatric	2516 (6.21)
General surgery	1490 (3.68)

[a]CNY = US $0.16.

Factors Associated With Patient Activeness During Online Medical Consultation

The correlation matrix of physician- and patient-level variables and their univariate effects on logPAS is shown in Table 2. No strong linear correlations existed between the predictor variables, given that all the Pearson correlation coefficients were between –0.1 and 0.5. In the bivariate models, the effects of almost all factors were significant (P<.001), except for physician age. Thus, the next step for multilevel regression included them all as predictors. Our model included the physician's age as a control variable, given that it was associated with the use of online medical consultation service in previous studies [22], although it failed to exert a significant effect in the univariate analysis (P=.45).

Table 2. Univariate regression result and correlation matrix for the variables of the study (N=40,505).

Variable	Univariate regression	Correlation matrix								
	logPAS[a]	1	2	3	4	5	6	7	8	9
1. Patient age	–0.00[b]	1								
2. Patient gender	0.04[b]	0.25[b]	1							
3. Patients' waiting time for response	–0.21[b]	–0.02[b]	0.03[b]	1						
4. Patients' initiation of consultation	0.92[b]	–0.03[b]	–0.02[b]	0.06[b]	1					
5. Patient cost for consultation	0.49[b]	0.02[b]	–0.04[b]	–0.08[b]	0.11[b]	1				
6. Physician age	0.00	0.07[b]	0.04[b]	0.12[b]	0.03[b]	0.23[b]	1			
7. Physician gender	0.30[b]	0.03[b]	0.26[b]	0.01[c]	0.03[b]	–0.02[c]	0.01	1		
8. Physician online consultation volume	–0.08[b]	–0.03[b]	0.06[b]	0.00	0.02[b]	0.30[b]	0.39[b]	0.07[b]	1	
9. Physician online consultation fee	0.01[b]	–0.06[b]	–0.05[b]	0.05[b]	0.06[b]	0.46[b]	0.34[b]	–0.06[b]	0.52[b]	1

[a]PAS: patient activeness score.

[b]P<.001.

[c]P<.01.

Multilevel Models

Two-level regression analyses using maximum likelihood estimation were conducted to model how patient- and physician-level factors were associated with patient activeness in online medical consultations. In Table 3, the full model includes first-level factors, such as patient demographics and consultation behavioral factors. Specifically, these consultation behavioral factors were considered explanatory variables with random slopes at the patient level, controlling for age and gender. To simplify random-slopes models, correlations between intercepts and slopes were removed by assuming that the random effects (intercepts and slopes) are independent. All first-level factors were significantly associated with patient activeness. Patients' waiting time for response (β=–.17; P<.001) during consultations showed a negative association with patient activeness, whereas patients' initiation of consultation (β=.83; P<.001) and patient cost for consultation (β=.52; P<.001) were positively associated with activeness.

In addition, all physician-related variables except age showed a substantial effect. In terms of gender, patients communicating with female (β=.09; P=.01) physicians scored higher on the activeness measure. Physicians' online consultation volume (β=–.10; P=.01) was negatively associated with patient activeness, whereas physician online consultation fee (β=.03; P=.01) was positively associated with patient activeness.

Also shown in Table 3 are the estimates from the full model with cross-level interactions between patients and physicians. The deviance difference test produced a chi-square of 15.5 (df=3; P=.001), indicating that the full model should be preferred compared with Model B without interactions (Multimedia Appendix 1). All the 3 proposed interactions were statistically significant at .05. The coefficient for the interaction between patients' waiting time for response and physician online consultation volume was 0.05. The coefficient for the interaction between patients' initiation of consultation and physician gender was –0.08, and the coefficients for the interaction between

XSL•FO
RenderX

patient cost for consultation service and physician online consultation fee was 0.03.

To better illustrate the interactive effect between patients' waiting time for response and physician online consultation volumes on patient activeness, physician online consultation volume was categorized into 3 categories: low, medium, and high. Figure 1 unveils that the negative relationship between patients' waiting time for response and patient activeness was negatively moderated by physicians' online consultation volumes (β=.05; *P*=.01). The longer patients waited for response, the less likely they were to actively participate in online medical consultation. In addition, patients tended to be even less active if their physicians had larger online consultations volumes than those with fewer volumes.

In the interaction between patients' initiation of consultation and physician gender, Figure 2 illustrates that the relationship between patients' initiation of consultation and patient activeness was negatively moderated by physician gender (β=–.08; *P*=.03). Patient activeness was substantially higher during online medical consultations initiated by patient themselves than by the physician. Particularly, when patients initiated the conversation, logPAS increased by 0.799 in the female physician condition and by 0.874 in the male physician condition. This interactive effect also suggests that when physicians initiated the conversation after the consultation has been launched, female physicians are likely to encourage patients to participate more actively than male physicians.

Figure 3 illustrates that the positive relationship between patient cost for consultation and patient activeness was positively moderated by physician online consultation fee (β=.03; *P*=.03). Patients who paid for their consultations were more likely to participate during online medical consultation. A higher physician online consultation fee was associated with more patient activeness from the paid patient.

Table 3. Multilevel models for patient activeness with individual- and physician-level factors (N=40,505 patients and 300 physicians).

Model	Full model
Variable	Coefficient (standard error)
Intercept	0.75 (0.10)[a]
Patient level	
Patient age	–0.00 (0.00)[a]
Patient gender	0.06 (0.01)[a]
Patients' waiting time for response	–0.17 (0.01)[a]
Patients' initiation of consultation	0.83 (0.02)[a]
Patient cost for consultation service	0.52 (0.03)[a]
Physician level	
Physician age	0.00 (0.00)
Physician gender	0.09 (0.04)[c]
Physician online consultation volume	–0.10 (0.04)[b]
Physician online consultation fee	0.03 (0.01)[c]
Cross-level interaction	
Patients' waiting time for response × physician online consultation volume	0.05 (0.02)[c]
Patients' initiation of consultation × physician gender	–0.08 (0.03)[c]
Patient cost for consultation service × physician online consultation fee	0.03 (0.01)[c]
Akaike information criterion	101,969.6
Deviance	101,933.6

[a]*P*<.001.
[b]*P*<.01.
[c]*P*<.05.

XSL•FO
RenderX

Figure 1. Interaction effects of patients' waiting time for response and physician online consultation volumes (POCVs) on patient activeness. Note: Mean POCV is the average value of physicians' online consultation volume. Higher and lower POCV were calculated by taking the mean POCV ± its SD.

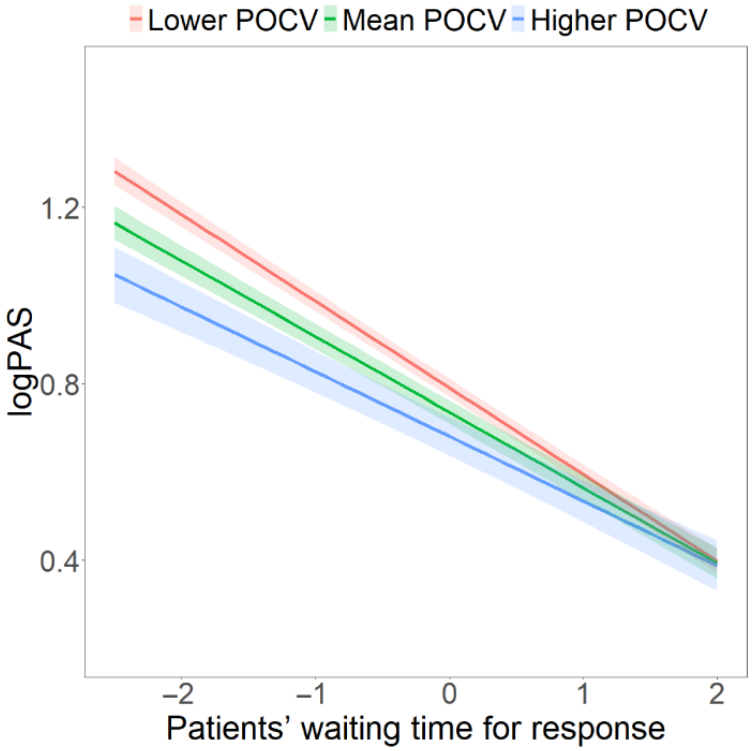
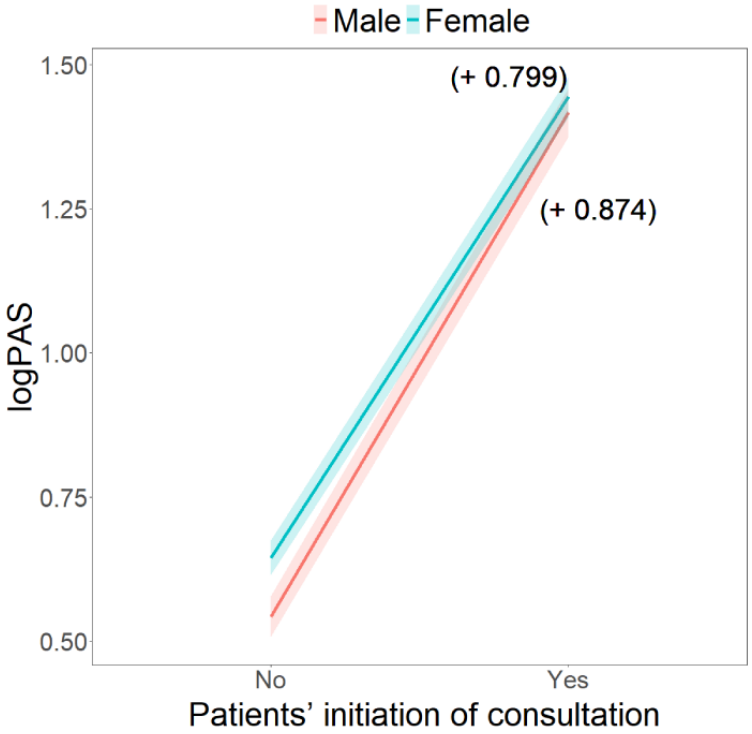
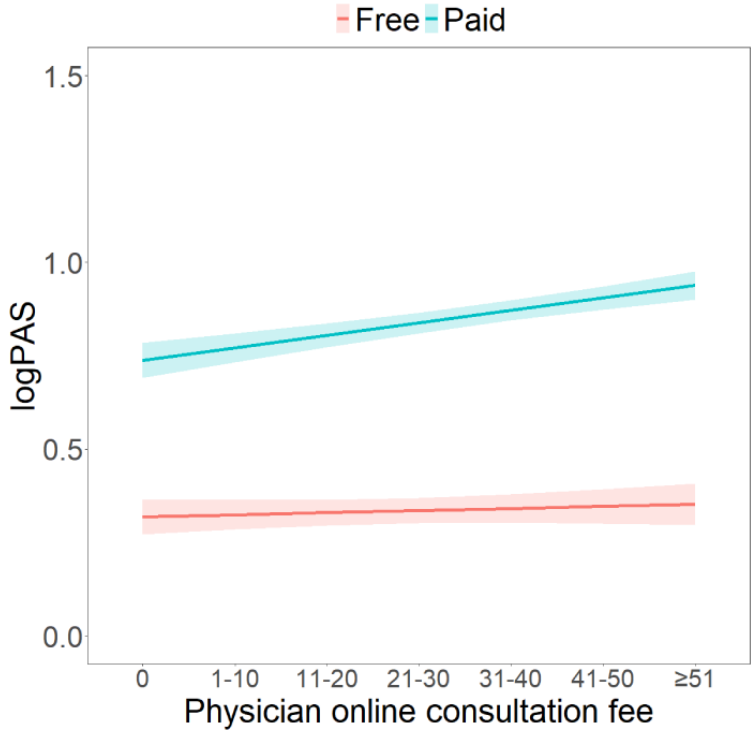


Figure 2. Interaction effects of patients' initiation of consultation and physician gender on patient activeness.



XSL•FO
RenderX

Figure 3. Interaction effects of patient cost for consultation and physician online consultation fee on patient activeness.



Discussion

Principal Findings

This study adopted a multilevel analysis to uncover factors associated with patient activeness in online medical consultation. Using a data set of 40,505 patients from 300 physicians in 10 specialties, this study found that patients were not equally active when participating in online medical consultation. Patient characteristics, consultation practices, and physician professional attributes are associated with patient activeness, such that (1) young and female patients tended to be more active in online medical consultation; (2) short waiting times, high cost, and patients' initiation of conversation were associated with higher patient activeness; (3) patients consulting with female physicians and physicians with relatively low online consultation volumes were more likely to actively participate in online medical consultation. The disparities in patient activeness do not only result from patient autonomy, but also from the interaction between physician professional practices and online consultation contexts. This study has implications for other low- and middle-income countries where medical resources are limited and hospital burdens are high.

This study found that the characteristics of the patients may be associated with their activeness. The age of the patients is associated with their activeness, echoing previous studies that found older patients were fearful or less confident in the "digital world" [23]. Older patients are more likely to believe and follow physicians' instructions, given that they often view doctors as responsible for medical decisions rather than themselves [24]. Furthermore, this study found that female patients were more likely to participate in online medical consultations. These results are consistent with other gender studies in the health care field, such that men are frequently underengaged with medical decision making due to masculinity concerns [23] and poor communication skills [25]. Nevertheless, we could not conclude that men and older patients are not necessarily incapable or unwilling to participate in online medical consultation. Given the advantages of patient activeness in shared medical decision making and treatment adherence [26], more men and older patients should be encouraged to actively participate during consultation.

In addition, this study suggests that patients' waiting time for response, initiation of conversation, and service cost were significant factors influencing patient activeness in online medical consultation. Previous studies have shown a negative relationship between waiting time and patient satisfaction in the offline settings [27], and the ability to save time is regarded as the most evident benefit for patients using online medical consultation services [28]. However, this study reveals that patients' median waiting time for an online physician response is approximately 2 hours in China, which is longer than most studies conducted in offline settings [27,29]. This challenges the naïve idea that online medical consultation typically saves time for patients. Given that most online medical consultation transpires in an asynchronous and discontinuous manner, online medical consultation could cost more time. Under this circumstance, it is important to recognize that responding timely becomes an even crucial factor for both patient activeness and patient satisfaction [1]. Moreover, patients' initiation of conversation in online medical consultation was positively associated with their activeness. In a traditional medical setting, physicians tend to be in the dominant position, initiating the conversation. Online communication grants patients the option to initiate the communication and indicate personal preferences. Patients who took the initiative to break the silence or make greetings demonstrate positive intentions for establishing rapport

XSL•FO
RenderX

with subsequent interactions, facilitating possible joint decision making [26]. Last but not least, cost for consultations is also associated with patient activeness. In contrast to free patients, patients who paid for consultation services were more likely to participate, given that their perceived input was higher and they expected better outcomes [30]. The patients who paid higher fees might be more motivated to actively participate in the interaction with physicians to maximize the high cost and make better medical decisions.

The study also shows that a physician's demographics and professional characteristics are associated with patient activeness. Patients who communicated with female doctors, instead of male doctors, were likely to be more active in online medical consultation. There have been several qualitative studies on how a physician's communication efforts can establish understanding and rapport with patients and encourage patient activeness [31]. Women physicians tend to have longer consultation time, engage in more partnership building, and are more interested in psychosocial aspects of health [25], allowing patients to become more engaged during consultation. Moreover, physicians with high online consultation volumes tended to discourage patients from participating more actively. It is a dilemma that many patients prefer to seek care from experienced chief physicians who have treated a large number of patients in China [32], but the chief physicians might not have enough time to interact with their specific patient because of the volume of patients. Patients lose their enthusiasm for active participation if they wait long for a physician with a large number of online consultations, as they may understand the physician's workload and expect less attention from him/her [33]. These findings validate the gender effect of physicians and suggest that the dynamics of patient-physician interactions are closely related to patient behaviors. Physicians are encouraged to play an integral part in increasing patient activeness during online consultation process.

Limitations

This study is not without limitations. Because of limited patient characteristics in our data set, we failed to identify more personal factors that may be associated with patient activeness. Furthermore, despite the high number of patients involved in this study, only 300 physicians from 10 different specialties were included. To verify current findings, future studies with a larger pool of physicians are required. In addition, text analysis can be synchronously conducted along with this study design to closely examine the interactive dynamics between patients and physicians during online medical consultation.

Conclusion

Relying on a multilevel analysis of 40,505 patients and 300 physicians, this study is among the early studies that identified a triangular model related to patient activeness in online medical consultation. The triangular factors include patient characteristics, consultation behavioral attributes, and physician professional characteristics. This study suggests that reducing patients' online waiting time and encouraging patients' initiation of consultation are related to the increase of patient activeness in low- and middle-income countries. The findings of this study have practical implications for expanding patient-centered services and improving patient experiences with online consultation services to reduce the pressure and burden of offline medical services.

Acknowledgments

This work has been supported by the National Social Science Fund of China (Grant/Award Numbers: 19ZDA324 and 20CXW024). We would like to thank Dr Taiquan Peng for his valuable suggestions for this study.

Conflicts of Interest

NL is the CEO and founder of Health 160 company.

Multimedia Appendix 1
Multilevel models for patient activeness with individual- and physician-level factors (N=40505, 300 physicians).
[DOCX File , 16 KB - jmir_v24i5e35557_app1.docx]

References

1. Wu H, Lu N. Online written consultation, telephone consultation and offline appointment: An examination of the channel effect in online health communities. Int J Med Inform 2017 Nov;107:107-119. [doi: 10.1016/j.ijmedinf.2017.08.009] [Medline: 29029686]
2. Jiang S. Talk to Your Doctors Online: An Internet-based Intervention in China. Health Commun 2021 Apr;36(4):405-411. [doi: 10.1080/10410236.2019.1692493] [Medline: 31736365]
3. Chen S, Guo X, Wu T, Ju X. Exploring the Online Doctor-Patient Interaction on Patient Satisfaction Based on Text Mining and Empirical Analysis. Information Processing & Management 2020 Sep;57(5):102253. [doi: 10.1016/j.ipm.2020.102253]
4. CGTN. Over 661 million Chinese tapped online for medical care in 2020. CGTN. 2021. URL: https://news.cgtn.com/news/2021-05-04/Over-661-million-Chinese-tapped-online-for-medical-care-in-2020-ZZr9PrWVl6/index.html [accessed 2022-04-18]

XSL•FO
RenderX

5. Jiang J, Cameron A, Yang M. Analysis of Massive Online Medical Consultation Service Data to Understand Physicians' Economic Return: Observational Data Mining Study. JMIR Med Inform 2020 Feb 18;8(2):e16765 [FREE Full text] [doi: 10.2196/16765] [Medline: 32069213]
6. Huang W, Cao B, Yang G, Luo N, Chao N. Turn to the Internet First? Using Online Medical Behavioral Data to Forecast COVID-19 Epidemic Trend. Inf Process Manag 2021 May;58(3):102486 [FREE Full text] [doi: 10.1016/j.ipm.2020.102486] [Medline: 33519039]
7. Jiang S. The Relationship between Face-to-Face and Online Patient-Provider Communication: Examining the Moderating Roles of Patient Trust and Patient Satisfaction. Health Commun 2020 Mar;35(3):341-349. [doi: 10.1080/10410236.2018.1563030] [Medline: 30821506]
8. Global Online Doctor Consultation Market Analysis. Global Market Estimate(GME). 2020. URL: https://www.globalmarketestimates.com/market-report/global-online-doctor-consultation-market-2172 [accessed 2021-04-18]
9. Cegala DJ. An exploration of factors promoting patient participation in primary care medical interviews. Health Commun 2011;26(5):427-436. [doi: 10.1080/10410236.2011.552482] [Medline: 21416422]
10. Ringdal M, Chaboyer W, Ulin K, Bucknall T, Oxelmark L. Patient preferences for participation in patient care and safety activities in hospitals. BMC Nurs 2017 Nov 21;16(1):69 [FREE Full text] [doi: 10.1186/s12912-017-0266-7] [Medline: 29200965]
11. Street RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
12. Jiang S. How Does Online Patient-Provider Communication Heal? Examining the Role of Patient Satisfaction and Communication Experience in China. Health Commun 2019 Nov;34(13):1637-1644. [doi: 10.1080/10410236.2018.1517634] [Medline: 30198772]
13. Lee SA, Zuercher RJ. A current review of doctor–patient computer-mediated communication. Journal of Communication in Healthcare 2017 Jan 31;10(1):22-30. [doi: 10.1080/17538068.2017.1282184]
14. Cegala DJ, Post DM. The impact of patients' participation on physicians' patient-centered communication. Patient Educ Couns 2009 Nov;77(2):202-208. [doi: 10.1016/j.pec.2009.03.025] [Medline: 19395225]
15. Chin J. Doctor-patient relationship: from medical paternalism to enhanced autonomy. Singapore Med J 2002 Mar;43(3):152-155. [Medline: 12005343]
16. Street RL, Millay B. Analyzing patient participation in medical encounters. Health Commun 2001;13(1):61-73. [doi: 10.1207/S15327027HC1301_06] [Medline: 11370924]
17. Wang Q, Adhikari S, Wu Y, Sunil T, Mao Y, Ye R, et al. Consultation length, process quality and diagnosis quality of primary care in rural China: A cross-sectional standardized patient study. Patient Educ Couns 2022 Apr;105(4):902-908 [FREE Full text] [doi: 10.1016/j.pec.2021.08.006] [Medline: 34391601]
18. Jin G, Zhao Y, Chen C, Wang W, Du J, Lu X. The length and content of general practice consultation in two urban districts of Beijing: a preliminary observation study. PLoS One 2015;10(8):e0135121 [FREE Full text] [doi: 10.1371/journal.pone.0135121] [Medline: 26258911]
19. Street RL, Gordon HS, Ward MM, Krupat E, Kravitz RL. Patient participation in medical consultations: why some patients are more involved than others. Med Care 2005 Oct;43(10):960-969. [doi: 10.1097/01.mlr.0000178172.40344.70] [Medline: 16166865]
20. Vahdat S, Hamzehgardeshi L, Hessam S, Hamzehgardeshi Z. Patient involvement in health care decision making: a review. Iran Red Crescent Med J 2014 Jan;16(1):e12454 [FREE Full text] [doi: 10.5812/ircmj.12454] [Medline: 24719703]
21. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
22. Hox JJ, Moerbeek M, van de Schoot R. Multilevel Analysis: Techniques and Applications, Third Edition. London, UK: Routledge; 2017.
23. Vaportzis E, Clausen M, Gow A. Older Adults Perceptions of Technology and Barriers to Interacting with Tablet Computers: A Focus Group Study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
24. Degner L, Sloan J. Decision making during serious illness: What role do patients really want to play? Journal of Clinical Epidemiology 1992 Sep;45(9):941-950. [doi: 10.1016/0895-4356(92)90110-9]
25. Street R. Gender differences in health care provider-patient communication: are they due to style, stereotypes, or accommodation? Patient Educ Couns 2002 Dec;48(3):201-206. [doi: 10.1016/s0738-3991(02)00171-4] [Medline: 12477604]
26. Poon BY, Shortell SM, Rodriguez HP. Patient Activation as a Pathway to Shared Decision-making for Adults with Diabetes or Cardiovascular Disease. J Gen Intern Med 2020 Mar;35(3):732-742 [FREE Full text] [doi: 10.1007/s11606-019-05351-6] [Medline: 31646455]
27. Alarcon-Ruiz CA, Heredia P, Taype-Rondan A. Association of waiting and consultation time with patient satisfaction: secondary-data analysis of a national survey in Peruvian ambulatory care facilities. BMC Health Serv Res 2019 Jul 01;19(1):439 [FREE Full text] [doi: 10.1186/s12913-019-4288-6] [Medline: 31262280]

XSL•FO
RenderX

28. Zanaboni P, Fagerlund A. Patients' use and experiences with e-consultation and other digital health services with their general practitioner in Norway: results from an online survey. BMJ Open 2020 Jun 17;10(6):e034773 [FREE Full text] [doi: 10.1136/bmjopen-2019-034773] [Medline: 32554721]
29. Ahmad BA, Khairatul K, Farnaza A. An assessment of patient waiting and consultation time in a primary healthcare clinic. Malays Fam Physician 2017;12(1):14-21 [FREE Full text] [Medline: 28503269]
30. Wu H, Lu N. Service provision, pricing, and patient satisfaction in online health communities. Int J Med Inform 2018 Feb;110:77-89. [doi: 10.1016/j.ijmedinf.2017.11.009] [Medline: 29331257]
31. Ruiz-Moral R. The role of physician-patient communication in promoting patient-participatory decision making. Health Expect 2010 Mar;13(1):33-44 [FREE Full text] [doi: 10.1111/j.1369-7625.2009.00578.x] [Medline: 19878341]
32. Burkle CM, Keegan MT. Popularity of internet physician rating sites and their apparent influence on patients' choices of physicians. BMC Health Serv Res 2015 Sep 26;15:416 [FREE Full text] [doi: 10.1186/s12913-015-1099-2] [Medline: 26410383]
33. Li Y, Yan X, Song X. Provision of Paid Web-Based Medical Consultation in China: Cross-Sectional Analysis of Data From a Medical Consultation Website. J Med Internet Res 2019 Jun 03;21(6):e12126 [FREE Full text] [doi: 10.2196/12126] [Medline: 31162129]

Abbreviations

ICC: intraclass correlation coefficient
PAS: patient activeness score

Edited by A Mavragani; submitted 09.12.21; peer-reviewed by J van Delden, U Bork, C Vorisek; comments to author 31.01.22; revised version received 12.03.22; accepted 09.04.22; published 27.05.22.

Please cite as:
Cao B, Huang W, Chao N, Yang G, Luo N
Patient Activeness During Online Medical Consultation in China: Multilevel Analysis
J Med Internet Res 2022;24(5):e35557
URL: https://www.jmir.org/2022/5/e35557
doi:10.2196/35557
PMID:35622403

©Bolin Cao, Wensen Huang, Naipeng Chao, Guang Yang, Ningzheng Luo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 27.05.2022. 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.

XSL•FO
RenderX

Original Paper

A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study

Géraldine Falgarone[1,2], MD, PhD; Guilhem Bousquet[1,3], MD, PhD; Arnaud Wilmet[4], MD; Albert Brizio[4], MD; Valérie Faure[4], MD; Celestin Guillouet[4], MD; Franck Baudino[4], MD; Isabelle Roque[5], MD; Samuel Mayol[6*], PhD; Frederic Pamoukdjian[1,7*], MD, PhD

[1]UMR_S942 MASCOT, INSERM, Université Sorbonne Paris Nord, Bobigny, France
[2]Unité de Médecine Ambulatoire, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[3]Service de Cancérologie, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[4]H4D (Health for Development), Paris, France
[5]Université Paris Cité, Paris, France
[6]Institut Universitaire de Technologie, Université Sorbonne Paris Nord, St-Denis, France
[7]Service de Médecine Gériatrique, Hôpital Avicenne, Assistance Publique–Hôpitaux de Paris, Bobigny, France
[*]these authors contributed equally

Corresponding Author:
Géraldine Falgarone, MD, PhD
Unité de Médecine Ambulatoire
Hôpital Avicenne
Assistance Publique–Hôpitaux de Paris
125 Rue de Stalingrad
Bobigny, 93009
France
Phone: 33 148026870
Fax: 33 148955258
Email: g.falgarone@aphp.fr

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/10/e43220/

Abstract

Background: Telemedicine technology is a growing field, especially in the context of the COVID-19 pandemic. Consult Station (Health for Development) is the first telemedicine device enabling completely remote medical consultations, including the concurrent collection of clinical parameters and videos.

Objective: Our aim was to collect data on the multisite urban and suburban implementation of the Consult Station for primary care and assess its contribution to health care pathways in areas with a low density of medical services.

Methods: In a proof-of-concept multisite prospective cohort study, 2134 consecutive patients had teleconsultations. Consultation characteristics were analyzed from both the patient and practitioner perspective.

Results: In this study, the main users of Consult Station were younger women consulting for low-severity seasonal infections. Interestingly, hypertension, diabetes, and preventive medical consultations were almost absent, while they accounted for almost 50% of consultations with a general practitioner (GP). We showed that for all regions where the Consult Station was implemented, the number of consultations increased as GP density decreased. The study of practitioner characteristics showed GPs from metropolitan areas are motivated to work with this device remotely, with a high level of technology acceptability.

Conclusions: The multisite implementation of Consult Station booths is suitable for primary care and could also address the challenge of "medical deserts." In addition, further studies should be performed to evaluate the possible contribution of Consult Station booths to limiting work absenteeism.

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e33507) doi:10.2196/33507

KEYWORDS

telemedicine booth; primary care; cost-benefit; absenteeism from work; health care system; telemedicine; consultation; remote medical consultation; proof-of-concept; general practice

Introduction

Alongside the development of the internet and connected tools over the past 2 decades, a rise in the development of eHealth technologies has been observed, facilitating remote communication between patients and caregivers [1,2]. This technological advancement meets the increasing need for more patient-centered medicine. Geographical, temporal, financial, cultural, and digital access issues are at the heart of these changes. Several digital communication systems and devices for telemedicine have been previously reported (eg, interactive voice response, SMS text messages, emails, interactive video, home-based videoconferencing, personal monitoring devices, and personal health records) [1,3-5].

However, telemedicine is not yet ubiquitous and there are ongoing debates on how to improve the quality of patient care. This is particularly true for teleconsultations [6,7]. Some general practitioners (GPs) remain skeptical of telemedicine, with emerging questions on cost-effectiveness, its impact on health outcomes and care, and its usefulness for people with chronic conditions or young, healthy people. For example, in France, the national health insurance has reimbursed teleconsultations since 2019 under specific conditions linked to the standard health care pathway for primary care and GPs [8], but its use was still limited in late 2019. Before the COVID-19 pandemic, many other barriers to adopting teleconsultations worldwide were identified, including barriers related to staff and programmers, patients (age and level of patient education), and practitioners (training, resources, type of device, ethics, confidentiality, and accountability) [6,9].

The year 2020 was seriously impacted by the global spread of COVID-19, which necessitated the promotion of new health care initiatives and a reorganization of telemedicine to meet patients' expectations for broader access [10-14]. The unstructured and opportunistic implementations of many telemedicine devices and protocols during the pandemic have cast light on the urgent need for standardization [15,16]. In France, the COVID-19 pandemic has accelerated the use of telemedicine, leading to better and wider reimbursement not only for GPs and specialists, but also for nurses, speech therapists, and midwives [17,18].

To date, none of the telemedicine technologies reported involve a single application that enables patients and physicians to conduct a comprehensive measurement of medical parameters. In 2009, Consult Station, a French telemedicine booth, was created and developed by Health for Development (H4D) to meet the growing needs of telehealth; it combines remote consultations, measurement of medical parameters, and diagnostic tools in a single location, and includes a dedicated training program for physicians.

In this proof-of-concept study, we report a multisite implementation of the Consult Station booth for primary care in France and its contribution to health care pathways in the context of generalization of telemedicine devices.

Methods

Study Design and Population

This was a multisite prospective observational cohort study that consecutively included all patients aged ≥18 years who had a teleconsultation via Consult Station in France from September 16, 2019, to January 31, 2020, with no exclusion criteria and no patient exclusion in the data analysis.

Ethics Approval

Informed consent was obtained from each patient before inclusion. Data extraction was anonymized. This noninterventional study obtained the approval of the local ethics committee for collecting and analyzing data (Avicenne hospital, number CLEA-2018-019; 020-019).

Description of Consult Station

H4D is a company specifically dedicated to clinical telemedicine [19] and it created the Consult Station booth in 2009. This European Class 2 certified autonomous medical device [20] has functions dedicated to the automated measurement of several medical parameters (weight, height, BMI, measures of pain, temperature, blood pressure, cardiac frequency, and oxygen saturation) and includes several diagnostic tools (pain scale, electrocardiogram, stethoscope, dermatoscope, audiometry, capillary glycemia, and otoscope). It has a video interface that enables remote consultations with a physician (Multimedia Appendix 1). There are two modes of teleconsultation: a self-performed checkup and a clinically assisted teleconsultation (deployed in this study). A team of 15 physicians was specifically trained on using the Consult Station booth before the booths were implemented. The physicians' training program was funded by H4D. Systematic cleansing, adapted to the COVID-19 pandemic, was performed by a trained technical agent between each patient. New booths are to include a UV-C lamp, which shortens the cleansing process to less than 3 minutes.

Access to Consult Station

Consult station booths were implemented on the premises of large companies and town halls, and employees were informed of the device's availability and told they had free access to it. When patients wanted a teleconsultation, they had to connect to an appointment booking website provided by H4D and agree to privacy and confidentiality rules. In accordance with the French law on teleconsultations, an appointment must be given to the patient within 48 hours. If necessary, a distant care manager helped the patient schedule the teleconsultation. There

XSL•FO
RenderX

were no restrictions on the use of the device and there was no need to be referred by a practitioner to book an appointment.

GP and Patient Characteristics

GPs were recruited on a voluntary basis and systematically trained. The GP characteristics collected for this study were age, gender, medical specialty, location of private practice, and time devoted to teleconsultations per week.

For each patient, data were collected by the physician during the teleconsultation. Data collected included age, gender, date, location of consultation (ie, Paris, Paris suburbs, or other regions), reasons for consultation, and classified consultation diagnosis according to the International Classification of Diseases, Tenth Revision (ICD-10).

Statistical Analysis and GP Density Indicators

Categorical data were expressed as numbers and proportions, while continuous data were expressed as mean (SD) or median (IQR) as appropriate.

The number of teleconsultations was assessed according to the local GP density per 100,000 inhabitants [21] and then according to the localized potential accessibility (LPA) to a GP for cities and rural administrative areas [22]. LPA is a composite indicator that considers both GP proximity and GP availability; it is the ratio of the number of completed consultations to the number of available consultations per inhabitant. An LPA value <2.5 per year is used by the French Ministry of Health to define the term “medical deserts” [23].

The data were analyzed and graphics were generated using R statistical software (version 4.0.0; R Foundation for Statistical Computing).

Results

Teleconsultation Characteristics

A total of 2134 teleconsultations were carried out from September 16, 2019, to January 31, 2020. The teleconsultations were distributed over weekdays as follows: 419 (20%) on Mondays, 450 (21%) on Tuesdays, 411 (19%) on Wednesdays, 454 (21%) on Thursdays, and 400 (19%) on Fridays. Medical parameters measured and diagnostic tools used were as follows: weight (344/2134, 16%), height (n=344, 16%), BMI (n=344, 16%), temperature (n=1450, 68%), blood pressure (n=1351, 63%), cardiac frequency (n=823, 38.5%), oxygen saturation (n=823, 38.5%), electrocardiogram (n=14, 0.6%), stethoscope (n=896, 42%), dermatoscope (n=156, 7%), and otoscope (n=924, 43%). A teleprescription was issued for 1567 (73%) patients. A sick leave certificate was issued for 42 (3%) patients. Complete data, including the reasons for teleconsultation, were available for 1746 (82%) patients. Overall, 98% (1715/1746) of the teleconsultations were conducted in full, while 2% (n=31) of teleconsultations were abandoned as a result of connection issues. Table 1 shows the distribution of the reasons for teleconsultation. Cough disorders, pain, joint diseases, and rhinitis were the most frequently provided reasons.

XSL•FO
RenderX

Table 1. Distribution of the reasons for teleconsultation among 1715 patients.

Reasons for teleconsultation	Patients, n (%)
Mild infectious diseases	
Cough disorders	343 (20)
Rhinitis	154 (9)
Fever, unspecified	137 (8)
Functional urinary symptoms	103 (6)
Pain	
Unspecified pains	187 (11)
Joint diseases/pain	137 (8)
Unspecified abdominal pain	51 (3)
Headache	51 (3)
Asthenia, skin, and allergy	
Asthenia	67 (4)
Skin disorders	51 (3)
Unspecified allergy	86 (5)
Prevention care and certificate	
Prescription renewal	51 (3)
Prevention	120 (7)
Laboratory results	343 (20)
Other[a]	154 (9)

[a]Other included unspecified visual disorders (n=19), gynecological disorders (n=17), unspecified vertigo (n=17), pregnancy (n=16), unspecified screening (n=15), nausea or vomiting (n=14), unspecified sleep disorders (n=8), myalgia (n=8), and psychological demands (n=7).

Use of Consult Station by Women

The main users of Consult Station were younger women with a mean age of 38.7 (SD 10.3; range 20-77) years. Table 2 shows the patient characteristics. The mean teleconsultation duration was 18 (SD 1.2) minutes. Overall, the diagnostic categories most often observed were otorhinolaryngology, osteoarticular pain, and routine clinical examinations, with no difference between women and the whole cohort. Prevention advice (vaccination, laboratory results, and addiction counseling) concerned only 2% (34/1715) of the patients. None of the patients consulted for hypertension- or diabetes-related follow-ups. Referral following a consultation did not occur for 58% (995/1715) of teleconsultations.

XSL•FO
RenderX

Table 2. Characteristics of 1715 consecutive patients with teleconsultations.

Variable	Whole cohort (N=1715), n (%)	Women (N=1230), n (%)
Age cohorts (years)		
20-39	948 (56)	722 (59)
40-59	723 (42)	488 (40)
≥60	34 (2)	20 (1)
Gender		
Women	1230 (72)	N/A[a]
Men	475 (28)	N/A
Diagnostic domains for teleconsultation		
Otorhinolaryngology	756 (44)	555 (45)
Osteoarticular	189 (11)	129 (11)
Normal clinical examination	187 (11)	111 (9)
Pneumonology	112 (7)	77 (6)
Dermatology	77 (5)	66 (6)
Urology	77 (5)	58 (5)
Gastroenterology	52 (3)	33 (3)
Ophthalmology	45 (3)	27 (2)
Abnormal laboratory results	35 (2)	27 (2)
Neurology	35 (2)	28 (2)
Prevention	34 (2)	25 (2)
Gynecology	29 (2)	24 (2)
Cardiovascular/high blood pressure	26 (2)	13 (1)
Psychiatry	18 (1)	15 (1)
Asthenia	12 (1)	9 (0.7)
Dental	10 (0.5)	8 (0.6)
Endocrinology/diabetes	6 (0.3)	4 (0.3)
Sexually transmitted infection	6 (0.3)	4 (0.3)
Missing data	8 (0.5)	7 (0.6)
Consultant recommendations		
No orientation	994 (58)	716 (58)
General practitioner	387 (23)	273 (22)
Complementary examination	159 (9)	115 (9)
Specialist	104 (6)	70 (6)
Other health professional	53 (3)	42 (3)
Emergency department	10 (0.5)	7 (0.6)
Missing data	8 (0.5)	7 (0.6)

[a]N/A: not applicable.

Consult Stations Were Mainly Deployed in Low-to-Moderate GP Density Areas

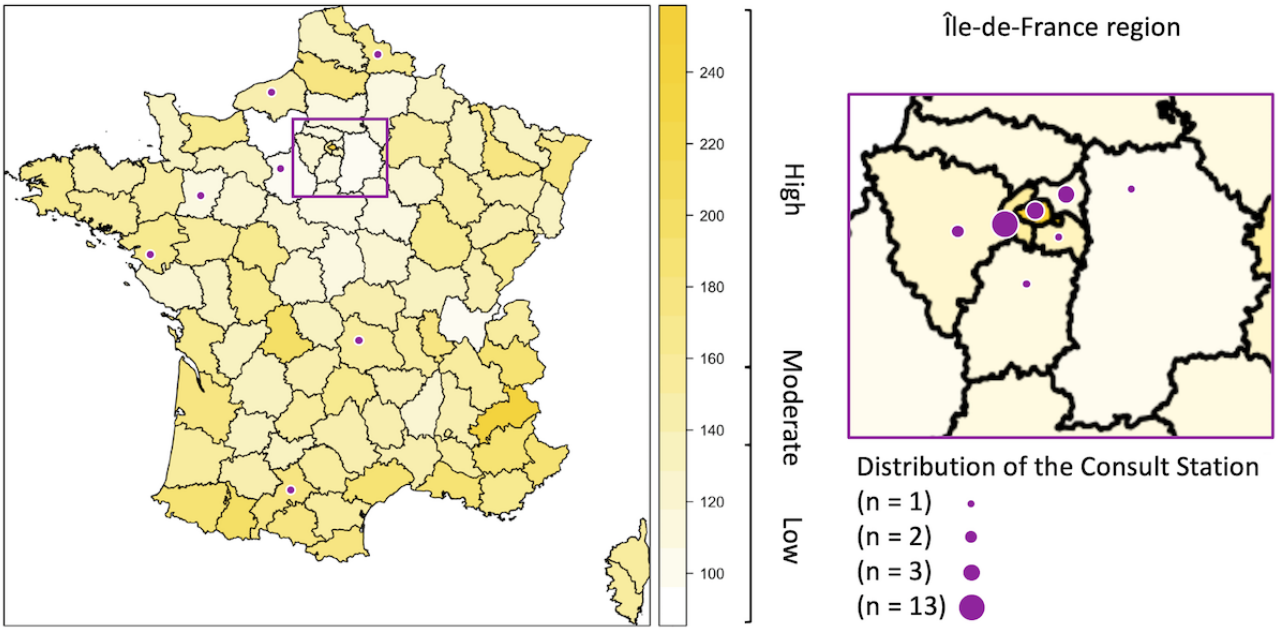
A total of 31 Consult Station booths were implemented in France for primary care management, mainly on the premises of large companies (≥5000 employees) and local authorities, with one of them set up inside a town hall (Figure 1). In the Île-de-France region (ie, Paris and its suburbs), 24 (77%) booths were implemented. The GP density of these areas ranged from 96/100,000 to 248/100,000 inhabitants (mean 149.7, SD 27). We classified GP density into 3 categories as follows: low density (96-137), moderate density (138-159), and high density (≥160). We observed that the Consult Station booths were

XSL•FO
RenderX

located mainly in moderate-density (16/31, 52%) or low-density areas (11/31, 35%).

We then considered a French composite indicator for access to a GP, namely LPA, which provides the completed number of GP consultations per patient in relation to the number of available GP consultations. Medical deserts are defined by an LPA value under 2.5 per year, which applies to 5.1% of France, while the national LPA value is 3.7 (range 1.4-12.1). Using this threshold of 2.5, none of the Consult Station booths were in a medical desert. We then further classified LPA into 3 categories as follows: low LPA (2.5-3.2), moderate LPA (3.3-4.0), and high LPA (≥4.1). This showed that 19% (6/31) and 55% (17/31) of the Consult Station booths were located in moderate- or low-LPA areas, respectively.

Figure 1. Implementation of Consult Station booths according to general practitioner density in France (left panel) and in the Île-de-France region (ie, Paris and its suburbs; right panel).



Consult Station Could Improve Access to Practitioners

Table 3 shows the number of teleconsultations recorded for 28 Consult Station booths according to local GP density and LPA.

The number of teleconsultations was high in the Paris suburbs where GP density is low (124 GPs/100,000 inhabitants) and the LPA value is moderate (3.3 consultations/year). Across France, the number of teleconsultations increased as GP density decreased (Figure 2A). In contrast, access to teleconsultation also increased as the LPA indicator increased (Figure 2B). This suggests that access to routine GP consultations was not a hindrance to the use of teleconsultations.

The mean age of the 15 GPs was 39 (SD 8.5, range 30-60) years and 10 (80%) GPs worked in high-LPA areas. The number of years since the GPs' graduation ranged from 3-35 years. Of the participants, 60% (9/15) worked in a mixed setting, in both private practice and a hospital, and 47% (7/15) worked in a group practice. None had been previously trained for teleconsultations, but 3 of them reported occasional experiences in teleconsultation. Reasons provided by the doctors for their choice to practice telemedicine included the following: the innovative aspect of this device, collaborative work, diversification of their activity, and provision of care to people in medical deserts. For 73% (11/15) of them, the COVID-19 pandemic had not influenced their perception of teleconsultation and 87% (13/15) would recommend teleconsultation to other colleagues. It is worth noting that they were urban practitioners, as none worked in a low-LPA area (Multimedia Appendix 2).

Table 3. Number of teleconsultations with the Consult Station according to general practitioner density and LPA (N=2134).

Area	Teleconsultations, n (%)	Mean general practitioner density[a]	C/D ratio[b]	Mean LPA[c]
Paris (center)	222 (10)	High (248)	0.9	High (4.5)
Other regions	660 (31)	Moderate (148)	4.4	High (4.6)
Paris suburbs	1252 (59)	Low (124)	10	Moderate (3.3)

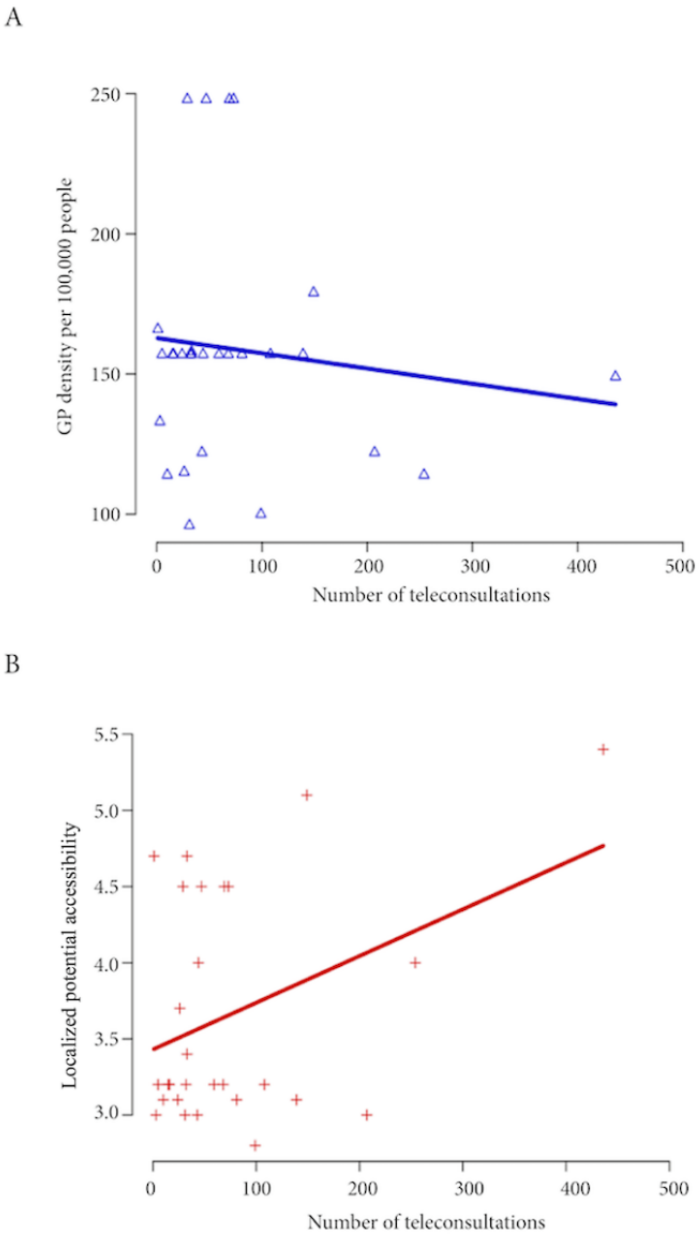
[a]General practitioner density in number per 100,000 inhabitants in France.

[b]C/D ratio: number of consultations/mean general practitioner density per 100,000 people.

[c]LPA: localized potential accessibility.

XSL•FO
RenderX

Figure 2. Scatter plot of the number of teleconsultations according to (A) GP density or (B) LPA. GP: general practitioner; LPA: localized potential accessibility.



Discussion

Principal Findings

The Consult Station booth is the first telemedicine device enabling completely remote medical teleconsultation with concurrent collection of clinical parameters, as otherwise teleconsultations are often limited to telephone consultations [24-26]. With real-time measurement of several medical parameters, the use of diagnostic tools, and video consultations, Consult Station is a good option when face-to-face consultations are not possible [26]. The COVID-19 pandemic contributed to an acceleration of teleconsultation acceptance and to the restructuring of pre-existing telehealth care devices or pathways. The Consult Station is a particularly original device appropriate for further health care standardization.

In our study, seasonal infections of low severity were the main reason for teleconsultations among younger patients. Interestingly, hypertension, diabetes, and preventive medical teleconsultations were almost absent, whereas they accounted for almost 50% of in-person consultations with a GP in France [27]. The Consult Station could offer a new, convenient health care pathway for younger patients with nonsevere health needs. Further studies are required to determine whether this new, convenient primary care pathway could help reduce visits to hospital emergency departments [28,29]. The results from our study could further the debate on the cost-effectiveness of telemedicine in wealthy countries. Interestingly, only 11% (188/1715) of the patients used the device for a routine clinical examination. Although our proof-of-concept study was not designed to determine whether teleconsultation leaned toward treatment of chronic diseases or more routine conditions, the Consult Station could be of great interest for systematic yearly checkups, particularly in areas with low GP density.

XSL•FO
RenderX

Furthermore, a recent study had shown that patients with chronic conditions are open-minded toward alternative modes of telemedicine [11], including their use for treating mental conditions [30].

Most of the patients were younger working women of childbearing age. This gender ratio might be explained by women being overrepresented in the use of the internet and telemedicine [11,15] and because time-saving is a major factor for telemedicine usage [31]. Appointments within 48 hours, convenient health care access comparable to private practice, and flexibility could all contribute considerably to patient satisfaction and acceptance of the Consult Station health care system, as reported with other telemedicine devices [32,33]. Furthermore, the system could help limit absenteeism from work due to illness if booths are implemented in the workplace as in our study [34,35]. In a recent study, the authors found that the rate of absenteeism from work was 3% among 5465 employees, with 56% of absences from women [35]. The rate of sick leave reached 28% overall and was 76% for younger women. In 2018, the annual mean cost of absenteeism from work was estimated at €4059 (US $4460) per individual in France [36], affecting 3.6% of employees. For a large company of at least 5000 employees, this would amount to a cost of €730,000 (US $802,198). In comparison, the minimum annual cost of a Consult Station booth would be €43,320 (US $47,604). This amount includes annual maintenance fees (€10,000, US $10,989), the annual cost equivalent to a full-time technical agent (€21,892, US $24,057) to clean the booth between each patient, and an amortization of the booth over 7 years (€11,428/year, total cost €80,000; US $12,558/year, total cost US $87,912). This could be an advantageous financial operation for companies to prevent work absenteeism. The question of work absenteeism should be addressed in a dedicated study including social and economic patient characteristics.

With the emergence of COVID-19, Consult Station could also be used to help manage patient flows in compliance with barrier measures [37,38].

With a multisite implementation, we believe that Consult Station booths could contribute to addressing the challenge of medical deserts. Even though they were largely implemented on business premises and none were in medical deserts, there was no real bias linked to the geographical distribution of Consult Station booths in our study, since 36% were implemented in areas with low GP density.

From the patients' perspective, the device offers easy access to doctors even in areas with low GP density. This implies a willingness among practitioners from metropolitan areas to respond to this challenge. Our study results showed a high level of technology acceptability among practitioners and our teleconsultation device addressed several of the barriers previously identified by GPs for the use of telemedicine. With acceptance by both patients and GPs, this type of teleconsultation device provides proof of concept for the generalization of telemedicine, and could succeed where public health policies have failed to address the growing problem of access to care in underpopulated rural areas [36]. Although our study was not designed to evaluate the impact of our device on vulnerable populations, we believe that it does not limit their health care access, as the health care system in France now enables reimbursement for teleconsultations for all patients.

Conclusions

The multisite implementation of Consult Station booths is suitable for primary care, but it also could meet the challenge of medical deserts. Although various types of telehealth or telemedicine facilities were already available in early 2020, the COVID-19 pandemic has highlighted the need for videoconsultations using remote tools such as those included in the Consult Station. In addition, further studies should be conducted to evaluate the possible contribution of Consult Station booths to limiting work absenteeism.

Acknowledgments

We thank Ms Angela Swaine and Ms Sarah Leyshon for English revisions. Health for Development (H4D) paid for the English language revision. We thank H4D for its authorization to reproduce a commercial image (ie, Multimedia Appendix 1) as part of this publication.

Authors' Contributions

GF, GB, AW, FB, and FP conceived and designed the study. AW, VF, AB, and CG collected the data. GF, GB, AW, FP, IR, and SM analyzed the data. GF, GB, AW, IR, and SM interpreted the data. GF, AB, AW, VF, CG, FB, FP, IR, and SM wrote and revised the manuscript. All authors read and approved the final manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

AW, VF, AB, CG, and FB are funded by Health for Development (H4D). GF, GB, IR, SM, and FP have no conflicts of interest to report.

Multimedia Appendix 1
The Consult Station booth.

XSL•FO
RenderX

[PNG File , 818 KB - jmir_v24i5e33507_app1.png]

Multimedia Appendix 2
Characteristics of the 15 general practitioners who performed teleconsultations.
[DOCX File , 15 KB - jmir_v24i5e33507_app2.docx]

References

1. Fortney JC, Burgess JF, Bosworth HB, Booth BM, Kaboli PJ. A re-conceptualization of access for 21st century healthcare. J Gen Intern Med 2011 Nov;26 Suppl 2:639-647 [FREE Full text] [doi: 10.1007/s11606-011-1806-6] [Medline: 21989616]
2. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07;9(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
3. Chambers R, Schmid M. Making technology-enabled health care work in general practice. Br J Gen Pract 2018 Mar;68(668):108-109 [FREE Full text] [doi: 10.3399/bjgp18X694877] [Medline: 29472202]
4. Mair F, Peterkin G, Laing A, Ferguson J, Fraser S. Feasibility of a telemedicine 'booth' for supporting remote care in Scotland. J Telemed Telecare 2008;14(3):127-129. [doi: 10.1258/jtt.2008.003007] [Medline: 18430277]
5. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. JMIR mHealth uHealth 2020 Oct 23;8(10):e18426 [FREE Full text] [doi: 10.2196/18426] [Medline: 33095175]
6. Iacobucci G. GP at Hand: NHS England upholds CCG's objection to planned Birmingham expansion. BMJ 2018 Sep 13;362:k3899. [doi: 10.1136/bmj.k3899] [Medline: 30213869]
7. O'Dowd A. Doctors question Hancock's idea of GP video consultations for all. BMJ 2018 Sep 14;362:k3934. [doi: 10.1136/bmj.k3934] [Medline: 30217932]
8. La téléconsultation. Ministère des Solidarités et de la Santé. 2021 Jul 26. URL: https://solidarites-sante.gouv.fr/soins-et-maladies/prises-en-charge-specialisees/telesante-pour-l-acces-de-tous-a-des-soins-a-distance/article/la-teleconsultation [accessed 2021-07-26]
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] [Medline: 29320966]
10. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
11. Oikonomidi T, Ravaud P, Barger D, Tran V. Preferences for alternative care modalities among French adults with chronic illness. JAMA Netw Open 2021 Dec 01;4(12):e2141233 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.41233] [Medline: 34964850]
12. Hasson SP, Waissengrin B, Shachar E, Hodruj M, Fayngor R, Brezis M, et al. Rapid implementation of telemedicine during the COVID-19 pandemic: perspectives and preferences of patients with cancer. Oncologist 2021 Apr;26(4):e679-e685 [FREE Full text] [doi: 10.1002/onco.13676] [Medline: 33453121]
13. Musaoğlu M, Yuksel M, Mizikoglu O, Arikan C. Telemedicine in monitoring pediatric LT patients before and during COVID-19 pandemic. Pediatr Transplant 2022 Feb;26(1):e14138 [FREE Full text] [doi: 10.1111/petr.14138] [Medline: 34505750]
14. Hall Dykgraaf S, Desborough J, de Toca L, Davis S, Roberts L, Munindradasa A, et al. "A decade's worth of work in a matter of days": The journey to telehealth for the whole population in Australia. Int J Med Inform 2021 Jul;151:104483 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104483] [Medline: 33984625]
15. Hamadi HY, Zhao M, Haley DR, Dunn A, Paryani S, Spaulding A. Medicare and telehealth: The impact of COVID-19 pandemic. J Eval Clin Pract 2022 Feb;28(1):43-48 [FREE Full text] [doi: 10.1111/jep.13634] [Medline: 34786796]
16. Saliba-Gustafsson EA, Miller-Kuhlmann R, Kling SMR, Garvert DW, Brown-Johnson CG, Lestoquoy AS, et al. Rapid implementation of video visits in neurology during COVID-19: mixed methods evaluation. J Med Internet Res 2020 Dec 09;22(12):e24328 [FREE Full text] [doi: 10.2196/24328] [Medline: 33245699]
17. Synthèses et bibliographies thématiques. IRDES. 2021. URL: http://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html [accessed 2022-03-16]
18. Téléconsultation et Covid-19: qui peut pratiquer à distance et comment? French Ministry of Health. 2021. URL: https://tinyurl.com/2t9kmp3u [accessed 2022-03-16]
19. Solutions de télémédecine clinique. H4D. URL: https://www.h4d.com/ [accessed 2022-03-16]
20. Council Directive 93/42/EEC of 14 June 1993 concerning medical devices. Official Journal of the European Communities. 1993 Jun 14. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31993L0042&from=FR [accessed 2022-03-16]
21. Professionnels de santé au 1er janvier 2018. INSEE. 2020. URL: https://www.insee.fr/fr/statistiques/2012677 [accessed 2022-03-16]

XSL•FO
RenderX

22. Accessibilité potentielle localisée (APL) aux médecins généralistes. Observatoire des Territoires. URL: https://www.observatoire-des-territoires.gouv.fr/accessibilite-potentielle-localisee-apl-aux-medecins-generalistes [accessed 2022-03-16]
23. Déserts médicaux: comment les définir? Comment les mesurer? Direction de la recherche, des études, de l'évaluation et des statistiques. 2017 May. URL: https://drees.solidarites-sante.gouv.fr/sites/default/files/2020-08/dd17.pdf [accessed 2021-01-29]
24. Banks J, Farr M, Salisbury C, Bernard E, Northstone K, Edwards H, et al. Use of an electronic consultation system in primary care: a qualitative interview study. Br J Gen Pract 2018 Jan;68(666):e1-e8 [FREE Full text] [doi: 10.3399/bjgp17X693509] [Medline: 29109115]
25. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method mtudy. J Med Internet Res 2018 Apr 17;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
26. Neeman E, Kumar D, Lyon L, Kolevska T, Reed M, Sundaresan T, et al. Attitudes and perceptions of multidisciplinary cancer care clinicians toward telehealth and secure messages. JAMA Netw Open 2021 Nov 01;4(11):e2133877 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.33877] [Medline: 34817586]
27. Quelques chiffres sur les résultats de consultation. Société Française de Médecine Générale. URL: https://tinyurl.com/3fws2vap [accessed 2022-03-16]
28. Pinchbeck EW. Convenient primary care and emergency hospital utilisation. J Health Econ 2019 Dec;68:102242 [FREE Full text] [doi: 10.1016/j.jhealeco.2019.102242] [Medline: 31605834]
29. Ugolini C, Leucci AC, Nobilio L, Bertè G. Reorganizing territorial healthcare to avoid inappropriate ED visits: does the spread of community health centres make walk-in-clinics redundant? BMC Health Serv Res 2020 Aug 27;20(1):807 [FREE Full text] [doi: 10.1186/s12913-020-05648-x] [Medline: 32854697]
30. Gentry MT, Puspitasari AJ, McKean AJ, Williams MD, Breitinger S, Geske JR, et al. Clinician satisfaction with rapid adoption and implementation of telehealth services during the COVID-19 pandemic. Telemed J E Health 2021 Dec 19;27(12):1385-1392. [doi: 10.1089/tmj.2020.0575] [Medline: 33606560]
31. Benis A, Banker M, Pinkasovich D, Kirin M, Yoshai B, Benchoam-Ravid R, et al. Reasons for utilizing telemedicine during and after the COVID-19 pandemic: an internet-based international study. J Clin Med 2021 Nov 25;10(23):5519 [FREE Full text] [doi: 10.3390/jcm10235519] [Medline: 34884221]
32. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
33. Polinski JM, Barker T, Gagliano N, Sussman A, Brennan TA, Shrank WH. Patients' satisfaction with and preference for telehealth visits. J Gen Intern Med 2016 Mar;31(3):269-275 [FREE Full text] [doi: 10.1007/s11606-015-3489-x] [Medline: 26269131]
34. Mat Saruan NA, Mohd Yusoff H, Mohd Fauzi MF, Wan Puteh SE, Muhamad Robat R. Unplanned absenteeism: the role of workplace and non-workplace stressors. Int J Environ Res Public Health 2020 Aug 24;17(17):6132 [FREE Full text] [doi: 10.3390/ijerph17176132] [Medline: 32846878]
35. Santa-Marinha MS, Teixeira LR, Godinho de Seixas Maciel EM, Ramos Moreira MDF. Epidemiological profile of sickness absenteeism at Oswaldo Cruz Foundation from 2012 through 2016. Rev Bras Med Trab 2018;16(4):457-465 [FREE Full text] [doi: 10.5327/Z1679443520180307] [Medline: 32754661]
36. Le cout caché de l'absentéisme. Institut Sapiens. 2018. URL: https://www.institutsapiens.fr/le-cout-cache-de-labsenteisme-au-travail-108-milliards-e-2/ [accessed 2018-11-21]
37. Armitage R. General practice after COVID-19: a greater role for remote patient monitoring. Br J Gen Pract 2021 Jun;71(707):276. [doi: 10.3399/bjgp21X716069] [Medline: 34045257]
38. Murphy M, Scott LJ, Salisbury C, Turner A, Scott A, Denholm R, et al. Implementation of remote consulting in UK primary care following the COVID-19 pandemic: a mixed-methods longitudinal study. Br J Gen Pract 2021;71(704):e166-e177 [FREE Full text] [doi: 10.3399/BJGP.2020.0948] [Medline: 33558332]

Abbreviations

GP: general practitioner
H4D: Health for Development
ICD-10: International Classification of Diseases, Tenth Revision
LPA: localized potential accessibility

XSL•FO
RenderX

Edited by A Mavragani; submitted 10.09.21; peer-reviewed by A Entezarjou, P Murphie, D Madhusudhan; comments to author 07.12.21; revised version received 26.01.22; accepted 05.02.22; published 17.05.22.

Please cite as:
Falgarone G, Bousquet G, Wilmet A, Brizio A, Faure V, Guillouet C, Baudino F, Roque I, Mayol S, Pamoukdjian F
A Teleconsultation Device, Consult Station, for Remote Primary Care: Multisite Prospective Cohort Study
J Med Internet Res 2022;24(5):e33507
URL: https://www.jmir.org/2022/5/e33507
doi:10.2196/33507
PMID:35254263

©Géraldine Falgarone, Guilhem Bousquet, Arnaud Wilmet, Albert Brizio, Valérie Faure, Celestin Guillouet, Franck Baudino, Isabelle Roque, Samuel Mayol, Frederic Pamoukdjian. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 17.05.2022. 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.

XSL•FO
RenderX

Original Paper

Online Video Teletherapy Treatment of Obsessive-Compulsive Disorder Using Exposure and Response Prevention: Clinical Outcomes From a Retrospective Longitudinal Observational Study

Jamie D Feusner[1,2,3,4], MD; Nicholas R Farrell[1], PhD; Jeremy Kreyling[1], BSc; Patrick B McGrath[1], PhD; Andreas Rhode[1], PhD; Ted Faneuff[1], MSW; Stephanie Lonsway[1], PhD; Reza Mohideen[1], BSc; John E Jurich[1], BA; Larry Trusky[1], BSc; Stephen M Smith[1], BA

[1]NOCD Inc, Chicago, IL, United States

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[3]General Adult Psychiatry & Health Systems Division, Centre for Addiction and Mental Health, Toronto, ON, Canada

[4]Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Jamie D Feusner, MD
NOCD Inc
225 N Michigan
Suite 1430
Chicago, IL, 60601
United States
Phone: 1 323 868 7211
Email: jamie@nocdhelp.com

Abstract

Background: Exposure and response prevention, a type of cognitive-behavioral therapy, is an effective first-line treatment for obsessive-compulsive disorder (OCD). Despite extensive evidence of the efficacy of exposure and response prevention (ERP) from clinical studies and in real-world samples, it is still underused as a treatment. This is likely due to the limits to access to care that include the availability of adequately trained therapists, as well as geographical location, time, and cost barriers. To address these, NOCD created a digital behavioral health treatment for OCD using ERP delivered via video teletherapy and with technology-assisted elements including app-based therapy tools and between-session therapist messaging.

Objective: We examined treatment outcomes in a large naturalistic sample of 3552 adults with a primary OCD diagnosis who received NOCD treatment.

Methods: The treatment model consisted of twice-weekly, live, face-to-face video teletherapy ERP for 3 weeks, followed by 6 weeks of once-weekly brief video teletherapy check-ins for 30 minutes. Assessments were conducted at baseline, at midpoint after completion of 3 weeks of twice-weekly sessions, and at the end of 6 weeks of brief check-ins (endpoint). Longitudinal assessments were also obtained at 3, 6, 9, and 12 months after endpoint.

Results: Treatment resulted in clinically and statistically significant improvements, with a 43.4% mean reduction in obsessive-compulsive symptoms (g=1.0; 95% CI 0.93 to 1.03) and a 62.9% response rate. Treatment also resulted in a 44.2% mean reduction in depression, a 47.8% mean reduction in anxiety, and a 37.3% mean reduction in stress symptoms. Quality of life improved by a mean of 22.7%. Reduction in OCD symptoms and response rates were similar for those with mild, moderate, or severe symptoms. The mean duration of treatment was 11.5 (SD 4.0) weeks, and the mean total therapist time was 10.6 (SD 1.1) hours. Improvements were maintained at 3, 6, 9, and 12 months.

Conclusions: In this sample, representing the largest reported treated cohort of patients with OCD to date, video teletherapy treatment demonstrated effectiveness in reducing obsessive-compulsive and comorbid symptoms and improved quality of life. Further, it achieved meaningful results in less than half the total therapist time compared with standard once-weekly outpatient treatment, an efficiency that represents substantial monetary and time savings. The effect size was large and similar to studies of in-person ERP. This technology-assisted remote treatment is readily accessible for patients, offering an advancement in the field in the dissemination of effective evidence-based care for OCD.

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e36431) doi:10.2196/36431

KEYWORDS

digital behavioral health; cognitive-behavioral therapy; CBT; exposure and ritual prevention; ERP; EX/RP; digital health; telehealth; cognitive therapy; obsessive compulsive disorder; OCD; clinical outcomes; teletherapy; remote therapy; telemedicine; obsessive compulsive; symptom; comorbid symptom; comorbidity; comorbidities; video therapy; virtual therapy; clinical outcome; patient outcome; online therapy; mobile health; mHealth; app based; health app; technology assisted; messaging

Introduction

Obsessive-compulsive disorder (OCD) is a prevalent and disabling psychiatric disorder, affecting 2.3% of individuals during their lifetimes [1]. Typically chronic if untreated, OCD is markedly detrimental to one's quality of life [2]. Yet, OCD can be treated effectively with psychotherapy or pharmacological interventions [3]. Exposure and response prevention (ERP), also known as exposure and ritual prevention, is a type of cognitive-behavioral therapy (CBT) that consistently demonstrates efficacy for OCD in numerous controlled trials and is also effective in less controlled clinical settings [3-7]. Based on this research evidence, ERP is considered a first-line treatment for OCD [8,9].

However, ERP requires specialty-trained therapists and thus is not readily available to everyone with OCD because of limited numbers of trained therapists, as well as cost and geographical limitations [10]. Indeed, the majority of individuals with OCD and related anxiety conditions are unable to access evidence-based psychotherapy [11]. Moreover, ERP typically requires over 25 hours of therapist time per patient [12] to achieve meaningful results; thus, when delivered in its most common format of once-weekly outpatient therapy, it could take 6 or more months.

To address the challenges of delivering ERP in terms of barriers to access and associated cost and time, NOCD has developed a digital behavioral health treatment program using video teletherapy. Remote ERP for OCD, delivered by video or telephone, has been demonstrated to significantly improve OCD symptoms [13]. Two head-to-head comparisons with in-person treatment in adults and adolescents show only small differences in outcome [14,15]. One of the several vital advantages of remote treatment is that therapists can readily interact with patients in the specific settings that most trigger their obsessional thoughts, images, or urges, for example in the home. This allows for administering in-session exposures that otherwise could be difficult or impossible to reproduce in an office setting. Although therapists in traditional face-to-face treatments can visit patients' homes and other nonoffice settings to administer exposures and help patients practice response prevention, this is logistically challenging and inefficient due to the travel times involved. Moreover, as of 2022, approximately 83% of the world's population (6.5 billion) owns a smartphone [16], and this grows yearly.

NOCD's treatment approach was inspired by a treatment previously tested in an open clinical trial [17] (N=33) that used the NOCD app integrated with brief in-person therapy. This trial tested a treatment protocol designed to minimize therapist time while increasing therapy intensity compared with once-weekly ERP sessions. It is possible that greater symptom reduction earlier in treatment, which may occur with more intensive treatment, could portend better ultimate clinical outcomes [18-20]. In a 2020 trial conducted by Gershkovich et al [17], there were high satisfaction ratings: 68.2% were "very" and 31.8% "mostly" satisfied with the services received. The treatment resulted in a mean reduction in OCD symptoms of 38.9%, with a response rate (≥35% reduction in OCD symptoms) of 52%. Mean therapist time was 6.7 (SD 1.52) hours total per patient.

We designed a treatment model for NOCD to treat patients with OCD using exposure and response prevention, with similar intensity, and to be able to reach as many as possible in the general community. To provide accessibility, all sessions were conducted remotely with video teletherapy. To provide additional support, enhance adherence, and potentially improve efficacy, every patient had access to between-session contact with their therapist via messaging. Further, a large online OCD community was available for further support through group message boards and scheduled support group sessions. In addition, peer support from individuals who had completed NOCD treatment was available to patients prior to starting treatment. The objective of this study was to examine treatment outcomes in a large naturalistic sample of 3552 adults with a primary OCD diagnosis who received NOCD treatment from January 1, 2020, to June 30, 2021.

Methods

Diagnostic Evaluations and Inclusion and Exclusion Criteria

Patients initially contacted the NOCD intake team as self-referrals or as referred from their health plans. They underwent diagnostic assessments by licensed clinical psychotherapists, who had received standardized training from NOCD in the evaluation and treatment of OCD using ERP. The diagnostic assessment consisted of a comprehensive clinical evaluation, including biopsychosocial elements of their history, and a standardized, semistructured diagnostic evaluation using the Diagnostic Interview for Anxiety, Mood, and Obsessive Compulsive and Related Neuropsychiatric Disorders (DIAMOND) [21]. Individuals who met DIAMOND criteria for OCD (consistent with Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [22]) as their primary disorder were treated. The majority of those who scored a 7 ("extreme") on the DIAMOND clinician-rated severity scale were referred to higher levels of care, including intensive outpatient programs, partial hospitalization programs, or residential treatment programs (exceptions were made on a case-by-case basis for a small number of individuals [n=16] whom the therapist and

XSL•FO
RenderX

clinical leadership deemed may benefit from treatment by NOCD). Other situations that resulted in referral at the time of diagnostic assessment included active substance use disorders or comorbid uncontrolled psychiatric disorders or symptoms (eg, mania, psychosis, or active suicidality), deemed to potentially interfere with treatment, on a case-by-case basis. Although the current analysis is of patients 18 years and older, NOCD treated those 5 years of age and older (results from the child and adolescent cohort are forthcoming). There was no upper limit on age. Medicated or unmedicated individuals were treated.

Treatment Model

The NOCD treatment model consisted of twice-weekly 60-minute remote ERP video sessions for 3 weeks. After this, patients had 6 weeks of once-weekly 30-minute video "check-in" sessions to guide ongoing ERP homework assignments conducted by the patients.

Therapists were trained and instructed to follow this framework for treatment but were allowed some flexibility to add sessions, if needed. In addition, between sessions, all patients had access to as-needed asynchronous text messaging with their therapists 5 out of 7 days per week to obtain guidance with exposures and response prevention. Patients had 24 hours per day and 7 days per week access to the online NOCD community, consisting of a forum of individuals around the world self-identified as having OCD, providing support and advice through online (monitored) postings. The NOCD app was available for patients to use during treatment; it provided tools for patients, in collaboration with their therapists, to create exposure hierarchies and do exposure exercises. Patients could also read and post messages in the NOCD community through the app.

All sessions were conducted via Zoom (US Health Insurance Portability and Accountability Act–compliant version). Patients could join the sessions via any personal computer or portable electronic device. For billing purposes, both the therapist and patient needed to be on video throughout the session. Aside from their electronic device, there was no other hardware required for either patients or therapists. Therapists were trained to not proceed with sessions if adequate sound and video quality could not be achieved, and in these scenarios, to reschedule in a timely manner. Additionally, during traditional daytime business hours (when most sessions were held) there was live technical support available to therapists to assist patients with troubleshooting if there were connectivity issues.

Therapists had Master's, PhD, or PsyD degrees, and were licensed in the states in which they provided remote treatment. Therapists received training by NOCD to conduct ERP and were provided ongoing group and individual supervision by experienced NOCD clinical leadership team members. All NOCD therapists received 3 days of intensive training on OCD, ERP, and application of ERP to OCD. After this training, there are several assessments that all clinicians must pass, including quizzes, a mock diagnostic session, a mock education session, and mock ERP sessions. As therapists go live, the clinical leadership team observes them in their first through fourth sessions randomly to see live examples of their diagnostic skills, provision of psychoeducation, and proficiency in the development of ERP hierarchies. The full-time therapists attend 2 hours per week of clinical supervision or case consultation as well as a 3-, 6-, 9-, and 12-month clinical advising review of their cases.

Assessments

Assessments were emailed to patients as links and were conducted at the initial diagnostic assessment, at treatment midpoint (after 6 twice-weekly therapy sessions), and the endpoint (after 6 weekly 30-minute check-in sessions). The use of patient-rated scales as the outcome variables of interest reduced the risk of therapist bias that may occur with clinician-rated scales. Follow-up assessments were sent to patients at the therapy visit closest in time to 3, 6, 9, and 12 months after their endpoint assessment. The majority of these follow-up sessions were 30-minute brief check-in sessions, as most had transitioned to less frequent visits (30-minute check-ins twice monthly to once every 3 months).

Dimensional Obsessive-Compulsive Scale (DOCS) [23] is a 20-item self-report measure of OCD symptom severity across four domains: contamination, responsibility for harm or mistakes, unacceptable thoughts, and incompleteness or symmetry. The DOCS has shown good psychometric properties, including strong convergent validity with the Yale-Brown Obsessive Compulsive Scale (r=0.54) and the Obsessive-Compulsive Inventory—Revised (r=0.69), and is sensitive to the effects of treatment.

The DIAMOND severity scale [21] is a 2-item clinician-rated assessment of the overall severity of an individual's emotional distress and functional impairment related to OCD symptoms. The clinician makes separate ratings of an individual's emotional distress and functional impairment on a scale ranging from 1 (Normal) to 7 (Extreme), and the higher of the two ratings is taken as the total severity score.

Depression, Anxiety, and Stress Scales (DASS-21) [24] is a 21-item self-report measure of symptoms of depression, anxiety, and stress. It has been widely used in previous research and has consistently shown good psychometric qualities.

Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form [25] is a 14-item self-report assessment of quality of life across a variety of life domains. It has demonstrated good psychometric properties in previous research.

Statistical Analyses

All data were deidentified prior to analysis. We analyzed data for those patients who completed at least the initial and the endpoint outcome assessments for the DOCS, the primary outcome measure. The majority also had a midpoint assessment. Data analysis was conducted using a linear mixed model (in part to handle missing data) with assessment time point as a fixed factor, patient as a random factor, and DOCS as the primary dependent variable. Secondary outcome analyses for the DASS-21 subscales of depression, anxiety, and stress, and the Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form, were analyzed using the same model. A tertiary outcome was follow-up symptom severity

XSL•FO
RenderX

ratings on the DOCS at 3, 6, 9, and 12 months from the endpoint assessment; this was also conducted using linear mixed models with assessment time point as a fixed factor (initial, 3-month, 6-month, 9-month, and 12-month time points), patient as a random factor, and DOCS as the primary dependent variable. Statistical significance was determined using an alpha of .05. Outcome analyses were conducted using SPSS version 27.0.0.0 (IBM Corp). We calculated Hedges g effect sizes using R (R Foundation for Statistical Computing).

Ethical Considerations

The analysis conducted in this study did not require research ethics board review as it does not meet the criteria for Human Subject Research as defined by federal regulations for human subject protections, 45 CFR 46.102(e); this is a secondary analysis of de-identified data from clinical records, obtained and analyzed retrospectively, and was not the result of a research intervention or interaction.

NOCD's Privacy Policy complies with the UK Data Protection Act of 2018, as well as the European Union's General Data Protection Regulation privacy law. All patients who are treated by NOCD must accept NOCD's Privacy Policy, which discusses how personal data are used, by whom, and for what purpose.

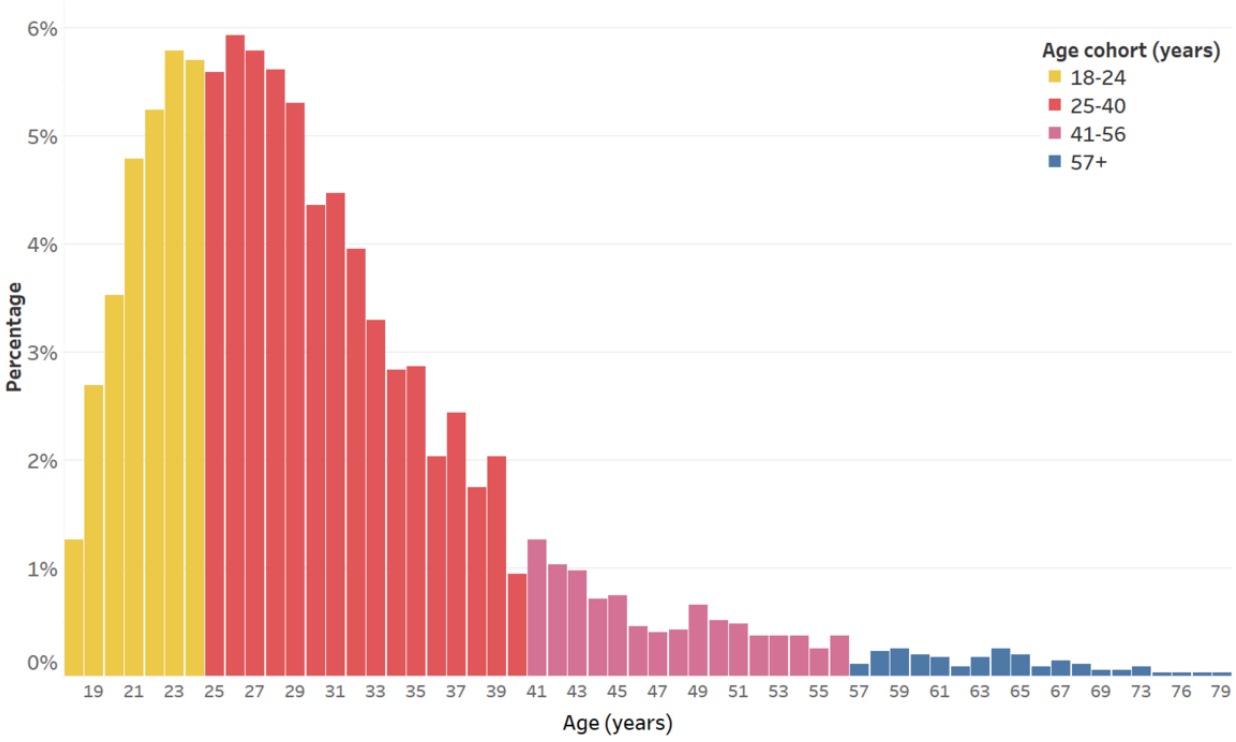
Results

Sample

We analyzed data collected from patients who started treatment between January 1, 2020, and June 30, 2021. (It is important to note that this date range was chosen to capture outcomes from when NOCD started enrolling substantial numbers of patients until the treatment protocol introduced minor changes in late August 2021; individuals who started as late as June 30, 2021, for example, would have finished treatment before these came into effect.)

We analyzed data from adults (aged ≥18 years) with a diagnosis of OCD who had at least an initial and endpoint assessment with the primary outcome measure, the DOCS. Data from 3552 patients who met these criteria were analyzed. Those who had fewer than 5 sessions were excluded (representing <0.1% of the sample), as this indicated that the treatment was likely interrupted, and outcomes were not available. The mean age was 29.9 (SD 9.3) years, range 18-79 years (Figure 1). In terms of gender, 55.88% (1985/3552) identified as female and 37.56% (1334/3552) identified as male (6.56% [233/3552] indicated nonbinary or another gender-expansive identity or did not provide this information). Regarding comorbidities, 36.4% (1293/3552) had a comorbid anxiety disorder, 32.8% (1165/3552) had a comorbid mood disorder, 10.3% (366/3552) had a comorbid OCD-related disorder, 5.3% (188/3552) had a trauma and stress-related disorder (posttraumatic stress disorder or acute stress disorder), 1.8% (64/3552) had a substance use disorder, 11.2% (398/3552) had another comorbid disorder, and 62% (2202/3552) had no comorbid disorders (Table S3 in Multimedia Appendix 1).

Figure 1. Age distribution.



XSL•FO
RenderX

NOCD App Use, Messaging, and NOCD Web-Based Community Posts

The app was used by 3529/3552 (99.4%) of patients at least once, and 3515/3552 (99%) sent at least one text message. Further, 1932/3552 (45.6%) made at least one community post. The mean number of app usages was 454.7 (SD 852.8), and the mean number of community posts was 55.8 (SD 282.0).

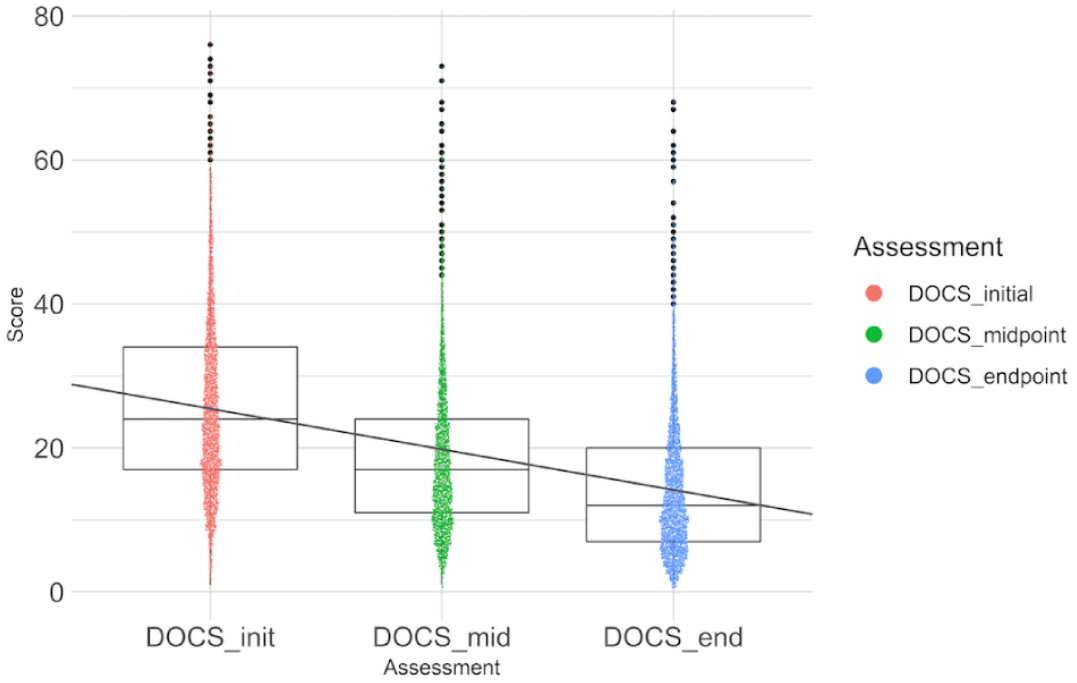
Treatment Duration

The mean treatment duration was 11.54 (SD 3.96) weeks (median=10.71, mode=9), the mean number of therapist sessions was 13.0 (SD 1.3; median=13.0, mode=13), and the mean number of therapist hours was 10.6 (SD 1.1; median=10.5, mode=10.5). Of the total 3552 sample, 53% (n=1883) had >14 sessions before the 3-month follow-up; of these, the mean number of 60-minute sessions was 7.7 (SD 2.0; the mean for those with 13 sessions was 7.0, SD 1.0) and the mean number of 30-minute check-in sessions was 9.1 (SD 1.7; the mean for those with 13 sessions was 5.8, SD 0.9). This amounts to a mean total of 16.8 (SD 2.2) sessions in those with >14 sessions; the majority of the additional sessions, if they were conducted, were check-in sessions.

OCD Symptom Results

NOCD treatment resulted in a significant decrease in patient-rated OCD symptoms (DOCS scores; $F_{6646.02}$=2810.08, P<.001; initial to endpoint Hedges g=1.0: "large" effect size). On the total sample level, DOCS scores improved from a mean of 26.0 (SD 12.3) to a mean of 14.7 (SD 9.8), representing a mean 11.3-point decrease (43.4%). On the individual patient level, the median DOCS score improvement was 45%. Note that we report the median for the individual score change rather than the mean, as it is a better representation of the central tendency for percentage change for these data. This is due to the fact that individuals' scores can worsen more than 100% but cannot improve more than 100%, which can result in a skewed distribution (Figure 2).

Figure 2. Changes in obsessive-compulsive disorder symptoms as assessed by the Dimensional Obsessive-Compulsive Scale (DOCS) with treatment (P<.001 for DOCS_mid compared with initial scores and P<.001 for DOCS_end compared with initial scores).



By midpoint, there were also statistically significant improvements; DOCS scores improved to a mean of 18.6 (SD 10.6), representing a mean 7.3-point decrease (28.2%). On the individual patient level, median DOCS score improvement was 30.8%.

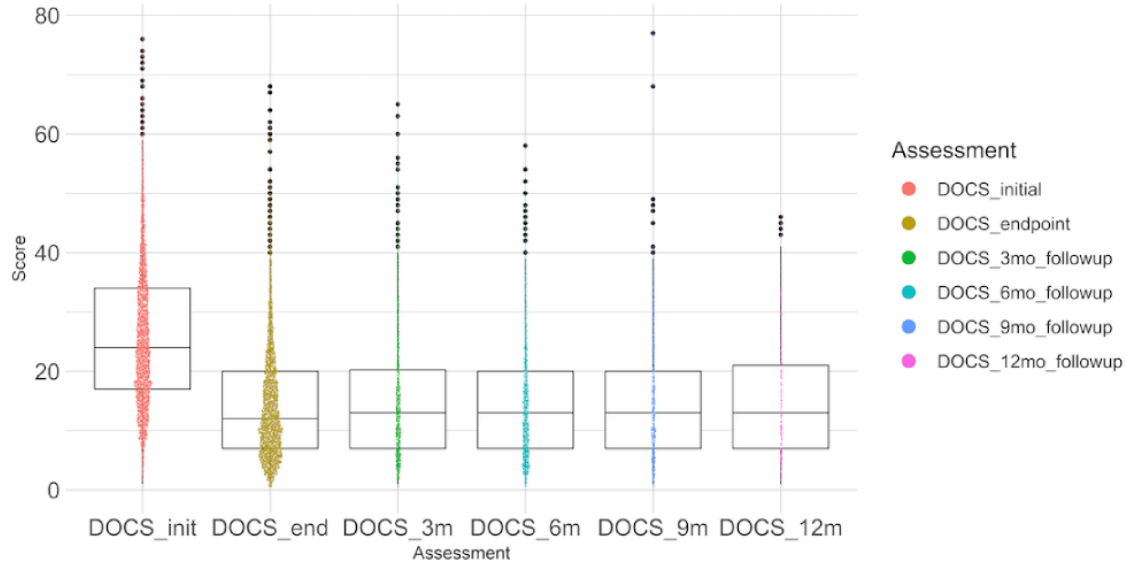
Further, 62.9% (2234/3552) met the criteria as full "responders," defined as a ≥35% reduction in OCD symptoms [26]. A total of 74.2% (2636/3552) met the criteria as achieving either partial (25%-35% reduction) or full response.

Follow-up Scores at 3, 6, 9, and 12 Months

Of the whole 3552 sample, 1633 (46%) did a 3-, 6-, 9-, or 12-month follow-up. At 3, 6, 9, and 12 months post the endpoint assessment, most patients had maintained their improvements in all symptom and quality of life domains. This was evidenced by mean DOCS, DASS depression, DASS anxiety, DASS stress, and QLESQ-SF scores at 3, 6, 9, and 12 months, which were similar to scores at the endpoint of treatment and remained significantly different from the initial assessment (Figure 3 and Tables S1 and S2 in Multimedia Appendix 1).

XSL•FO
RenderX

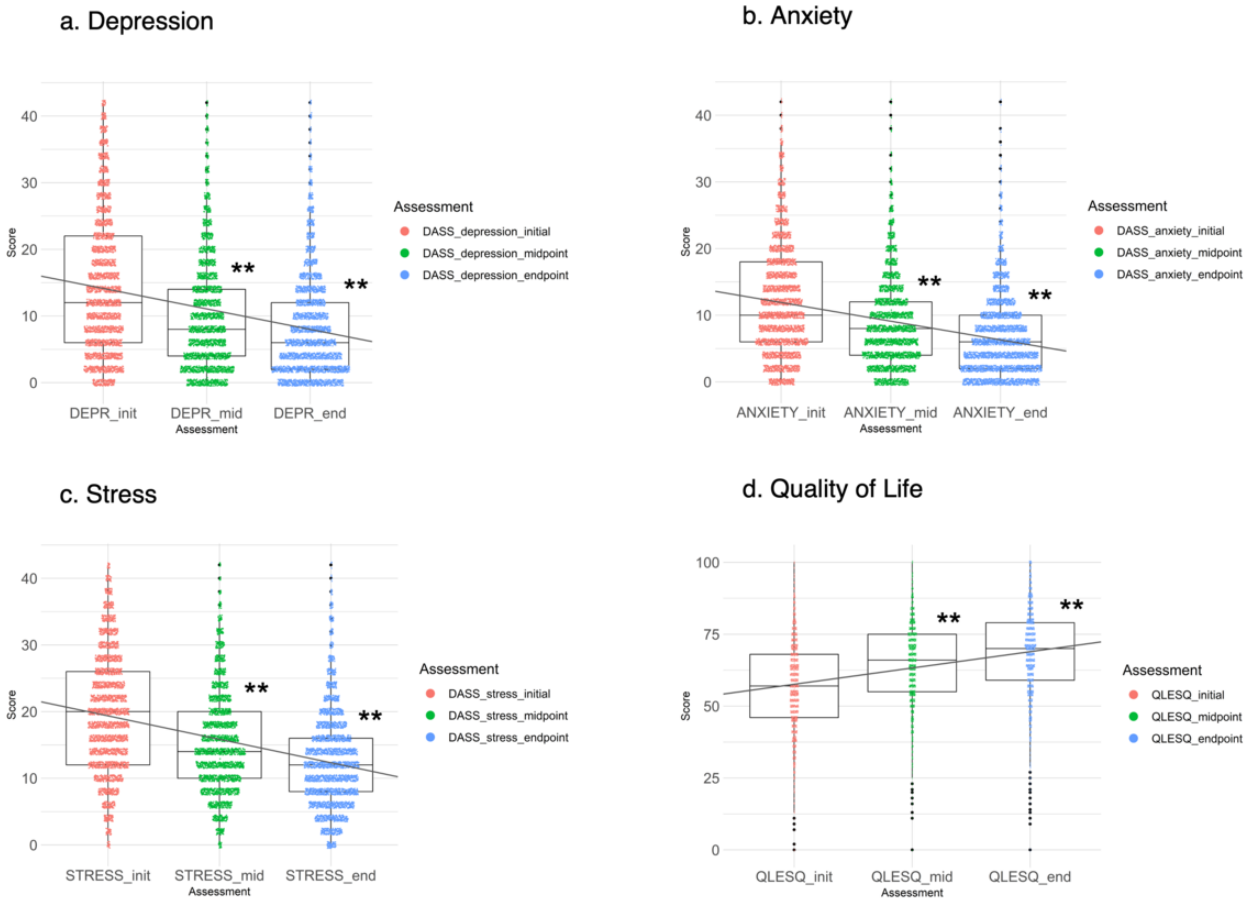
Figure 3. Longitudinal follow-up of obsessive-compulsive disorder symptoms as assessed by the Dimensional Obsessive-Compulsive Scale (DOCS; *P*<.001 for DOCS_end, DOCS_3m, DOCS_6m, DOCS_9m, and DOCS_12m compared with initial scores).



Depression, Anxiety, Stress, and Quality of Life Results

Treatment resulted in significant improvements on the DASS depression ($F_{6647.79}$=972.91, *P*<.001; initial to endpoint Hedges g=0.66), DASS anxiety ($F_{6659.83}$=1162.76, *P*<.001; initial to endpoint Hedges g=0.76), DASS stress ($F_{6645.12}$=1387.22, *P*<.001; initial to endpoint Hedges g=0.87), and the QLESQ-SF ($F_{6156.13}$=1140.66, *P*<.001; initial to endpoint Hedges g=0.76) (Figure 4 and Tables 1 and 2).

Figure 4. Changes in depression, anxiety, stress, and quality of life with treatment. DASS: Depression, Anxiety, and Stress Scales; DEPR: depression; QLESQ: Quality of Life Enjoyment and Satisfaction Questionnaire; ***P*<.001 compared with initial scores.



XSL•FO
RenderX

Table 1. Clinical assessments by treatment time point.

Outcome scale and assessment time point	Valid, n	Missing, n	Mean	SD	Mean 95.0% CI, lower bound	Mean 95.0% CI, upper bound	Median	Median 95.0% CI, lower bound	Median 95.0% CI, upper bound
DOCS[a]									
Initial	3552	0	26.0	12.3	25.5	26.4	24	24	25
Midpoint	3037	515	18.6	10.6	18.2	19.0	17	17	18
Endpoint	3552	0	14.7	9.8	14.4	15.0	12	12	13
DASS[b] depression									
Initial	3551	1	14.5	10.4	14.1	14.8	12	12	14
Midpoint	3032	520	10.2	8.7	9.9	10.5	8	8	10
Endpoint	3526	26	8.4	7.9	8.1	8.6	6	6	8
DASS anxiety									
Initial	3551	1	12.1	8.3	11.8	12.4	10	10	12
Midpoint	3033	519	8.6	6.7	8.4	8.8	8	8	10
Endpoint	3528	24	6.5	5.9	6.3	6.7	6	6	8
DASS stress									
Initial	3550	2	19.7	8.8	19.	19.9	20	20	22
Midpoint	3033	519	15.1	7.7	14.9	15.4	14	14	16
Endpoint	3528	24	12.6	7.3	12.4	12.9	12	12	14
QLESQ[c]									
Initial	3469	83	57.1	16.2	56.6	57.7	57	57	59
Midpoint	2764	788	64.5	14.9	63.9	65.0	66	66	68
Endpoint	3295	257	68.4	15.1	67.9	69.0	70	70	71

[a]DOCS: Dimensional Obsessive-Compulsive Scale.

[b]DASS: Depression Anxiety and Stress Scale—21.

[c]QLESQ: Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form.

XSL•FO
RenderX

Table 2. Changes in OCD,[a] depression, anxiety, and stress symptoms and quality of life by assessment time point.

Outcome scale and assessment	Score change	SE	Score change 95% CI, lower bound	Score change 95% CI, upper bound	Percentage change	*df*	*t*	Sig.[b]	Hedges g effect size	Hedges g 95% CI, lower bound	Hedges g 95% CI, upper bound
DOCS[c]											
Midpoint	–7.4	0.2	–7.7	–7.0	–28.4	6680.86	–45.71	<.001	0.66	0.63	0.69
Endpoint	–11.3	0.1	–11.6	–11.0	–43.4	6583.44	–73.97	<.001	1.00	0.93	1.03
DASS[d] depression											
Midpoint	–4.4	0.1	–4.7	–4.1	–30.4	6696.36	–29.16	<.001	0.48	0.45	0.51
Endpoint	–6.2	0.1	–6.4	–5.9	–42.5	6571.51	–42.96	<.001	0.66	0.62	0.69
DASS anxiety											
Midpoint	–3.6	0.1	–3.9	–3.4	–29.9	6715.28	–29.11	<.001	0.50	0.46	0.53
Endpoint	–5.6	0.1	–5.9	–5.4	–46.4	6572.55	–47.65	<.001	0.76	0.72	0.80
DASS stress											
Midpoint	–4.6	0.1	–4.9	–4.3	–23.4	6702.45	–32.23	<.001	0.59	0.55	0.63
Endpoint	–7.0	0.1	–7.3	–6.8	–35.8	6554.35	–51.96	<.001	0.87	0.83	0.91
QLESQ[e]											
Midpoint	7.7	0.3	7.2	8.2	13.5	6227.25	29.47	<.001	0.55	0.52	0.59
Endpoint	11.6	0.2	11.1	12.1	20.3	6107.25	46.98	<.001	0.76	0.72	0.79

[a]OCD: obsessive-compulsive disorder.

[b]Sig.: significance probability.

[c]DOCS: Dimensional Obsessive-Compulsive Scale.

[d]DASS: Depression Anxiety and Stress Scale–21.

[e]QLESQ: Quality of Life Enjoyment and Satisfaction Questionnaire—Short Form.

Post Hoc Analysis of Outcomes Stratified by Starting Clinician-Rated Severity Level

To determine how treatment response differed by different initial severity levels of OCD, we used the DIAMOND scale at the initial assessment to stratify patients into three groups of severity ratings: “Mild” (severity score of 2 or 3), “Moderate” (severity score of 4 or 5), or “Severe” (severity score of 6 or 7). Moreover, of the 3552 patients, 596 (17%) were missing DIAMOND severity scale scores. Of these, there was a median 46.86% reduction in DOCS scores and a 64.3% response rate. For DOCS scores, on the individual patient level, the Mild group (n=679, 19%) had a median 50.0% reduction, the Moderate group (n=2079, 59%) a median 42.9% reduction, and the Severe group (n=198, 6%) a median 44.6% reduction. Response rates from the DOCS were 68.8% for Mild, 60.7% for Moderate, and 61.6% for Severe.

Discussion

Patients with OCD treated with digital teletherapy using ERP show significant improvement in symptoms. OCD symptoms were reduced by 43.4%. Moreover, 62.9% (2234/3552) were classified as full responders, and 74.2% (2636/3552) had partial or full response. Treatment also resulted in improvements in the common comorbid symptoms of depression, anxiety, and stress and resulted in a significant improvement in quality of life. This provides evidence that a single, focused OCD treatment can result in an overall reduction of multiple disabling and distressing symptoms and improve the lives of patients. This is notable considering the fact that OCD is a chronic illness that individuals on average have for 11 years before receiving treatment [27]. Long-term follow-up data at 3-, 6-, 9-, and 12 months post treatment showed overall maintenance of gains from the initial treatment period.

These results demonstrate not only the magnitude of the effect of this treatment model on OCD and comorbid symptoms but also its efficiency in terms of cost and time savings. The time frame of these improvements was less than 12 weeks and less than 11 total therapist hours, on average. This is less than half the total therapist time and less than half of the duration of traditional once-weekly outpatient ERP [12]. This has the potential for substantial cost savings for patients and third-party payors such as health insurers.

This treatment format was inspired by a treatment previously developed and tested [17] to provide evidence-based ERP treatment for OCD, in a manner that is efficient in terms of total therapist time. The OCD symptom reduction results in the current NOCD-treated sample are similar to those achieved in that study. Yet, direct comparisons are limited by the fact that the current sample was from a “real-world” clinical setting rather than a controlled research setting with more selective inclusion

XSL•FO
RenderX

and exclusion criteria. Other differences that preclude direct comparisons include, but are not limited to, the fact that NOCD used the patient-rated DOCS scale, whereas the previous study used the clinician-rated Yale-Brown Obsessive-Compulsive Scale [28] as the primary outcome measure. Further, NOCD treatment consisted of face-to-face teletherapy rather than in-person therapy.

There are other important additional elements of treatment in the NOCD model that impact patient experience and may have influenced outcomes. Additional support for patients was available between sessions through patient-therapist SMS messaging. Patients also had 24-hour access to NOCD's web-based support community, consisting of messaging boards from others with OCD and organized around common OCD subtypes. This allows people to find others who experience similar symptoms, which can help reduce the sense that their OCD symptoms are a rare or unique type of OCD and therefore difficult to treat or may not even be OCD. This can be an important experience for those with OCD, given the broad and heterogeneous content of obsessional thoughts [29]. In addition, only a limited number of subtypes such as those involving contamination or washing, checking symptoms, and ordering or symmetry are typically described in the literature and are widely known, so certain OCD symptoms might be missed or misidentified by clinicians, family members, and patients themselves. In addition, patients had peer support from former patients who had completed NOCD treatment. When used, this would occur in the interval between contacting the initial call center for NOCD and their first diagnostic appointment with their therapist. The peer support may encourage people to follow through with scheduling and attending their first assessment meeting and beginning treatment. This additional support may be particularly useful due to the fact that ERP can be challenging for individuals to engage in; this is because, by necessity, ERP's therapeutic mechanisms are predicated on inducing distress (exposures) and eliminating behaviors that temporarily relieve distress (response prevention) but that perpetuate the cycle of obsessions and compulsions. Future investigations will quantify if, and to what degree, these additional digital and personal treatment elements affected clinical outcomes and patient experience.

Technology assistance likely played an important role in this treatment's ability to both engage and treat a large number of patients in wide-ranging geographic locations and to achieve a high mean rate of symptom improvement and a high rate of treatment response. Teletherapy using video allows people in remote locations to access treatment and to be able to complete, in-session, in vivo exercises in places and situations that are most relevant to, or triggering of, their symptoms. Previous studies of remote therapy demonstrate effect sizes that are similar to controlled studies of in-person treatment (see the meta-analysis [13]). The effect size for OCD symptom severity reduction in the current analysis of g=1.0 ("large" effect size) (95% CI .93 to 1.03) is similar to that found in a recent meta-analysis of controlled studies of ERP vs psychological placebo (g=1.13, 95% CI 0.71 to 1.55; 10 studies) [5]. Importantly, the current results are observed in a cohort that is one to two orders of magnitude larger than previous controlled ERP studies [5], providing strong evidence that virtual face-to-face ERP can be at least as effective as in-person ERP.

While most previous studies of ERP *efficacy* have come from clinical research trials, a meta-analysis of *effectiveness* studies of CBT in real-world clinical settings found an effect size for reduction of OCD symptoms across 11 studies of d=1.32 (95% CI: 1.19 to 1.45) [7]. A study published more recently examined CBT outcomes in an outpatient setting and found a mean 47.09% reduction in OCD symptoms on the Obsessive-Compulsive Inventory—Revised (n=451 at baseline and n=235 post treatment, effect size d=1.18) [30]. However, some differences limit direct comparisons to the current results because other studies used different OCD outcome rating scales (primarily the clinician-rated Yale-Brown Obsessive-Compulsive Scale or the patient-rated Obsessive-Compulsive Inventory—Revised rather than the DOCS) and had much smaller sample sizes.

Aside from video teletherapy, there are other technology-based features of this treatment that may have enhanced patient engagement. This includes integrated SMS messaging that allowed for increased continuity of treatment; patients could obtain advice and assistance when doing homework assignments in between sessions or when encountering unexpected situations that lead to obsessions and distress. This both helps keep treatment momentum and helps patients feel a more continuous sense of support. Further, the NOCD app has built-in tools for creating ERP hierarchies for exposure treatment planning. In addition, there are tools such as distress ratings to track progress during exposures and to track exposure-to-exposure progress, all of which can be visualized graphically by the patients and therapists. In this sample, almost all used the NOCD app and almost half made at least one post in the online community, with an average of approximately 56 posts per person. The specific effects of these technology features, as well as the effects of peer support and online community support, can be measured and evaluated in future analyses.

Another finding of note in this analysis was that symptom improvements were relatively similar for those with mild, moderate, and severe OCD symptoms. Overall mean symptom improvements were thus not driven only by those, for example, on the milder end of the symptom severity spectrum. Rather, the treatment model works well for those with a wide range of baseline symptoms, including those with severe OCD.

There are several limitations of this analysis, which are mostly due to its observational nature. Data were missing for some patients for certain rating scales (Table 1). Although all therapists received training in conducting ERP from NOCD's curriculum and learned the overall structure of the treatment model, therapy sessions were not videotaped to ensure treatment fidelity and consistency from therapist to therapist, as in a research study. However, therapists were regularly audited in terms of outcomes, patient feedback, and patient retention and were assisted in improving in any of these areas if necessary. Another limitation to the generalizability of the results is that the treatment model allowed for some flexibility; for example, therapists sometimes extended treatment beyond the 3 weeks of twice-weekly therapy or 6 weeks of once-weekly brief therapy check-ins if they deemed it important for patient improvement.

XSL•FO
RenderX

Another limitation is the use of the DOCS as the primary OCD outcome measure. As a patient-rated measure, it depends on patients' understanding of their symptoms in the framework of OCD. This could be problematic if patients do not recognize that some of their experiences are OCD symptoms, or if they believe that some experiences are OCD symptoms when they are not, which can result in erroneously low or high scores, respectively. This is a limitation, however, of all OCD rating scales to varying degrees. In addition, the majority of those whom therapists determined in the initial diagnostic assessment to have "extreme" OCD symptoms on the DIAMOND severity scale (aside from n=16 for whom exceptions were made on a case-by-case basis) were not treated and were instead referred to a higher level of care. Thus, although outcomes were similar for those with mild to severe cases, there is less certainty about generalization to those with extreme OCD severity. Another limitation is that 3-, 6-, 9-, and 12-month follow-up data were not available for many who completed the treatment. Thus, it remains unknown whether the proportions of those for whom 3-, 6-, 9-, and 12-months follow-up data were not provided represented individuals who were doing worse and sought other treatment, or were doing much better and did not see the need to continue these sessions. Further, even for those who provided data during this follow-up period, some may have engaged in other concurrent treatments.

In sum, ERP delivered in a technology-assisted video teletherapy treatment format results in clinically significant symptom and quality of life improvements in a real-world sample, on a large scale. This can provide a readily accessible means of obtaining effective, evidence-based treatment of OCD. Further, the relatively efficient treatment that is delivered can represent substantial cost savings for patients and third-party payors over traditional weekly outpatient face-to-face ERP.

Acknowledgments

The authors would like to thank the NOCD Therapists, Care Team, and Peer Advisors for their help in facilitating treatment and care experiences. There was no external funding for the analysis.

Conflicts of Interest

JDF, NRF, JK, PBM, AR, TF, SL, JEJ, LT, and SMS report personal fees from NOCD Inc during the conduct of the study.

Multimedia Appendix 1
Supplemental document that contains visualizations of analysis pertaining to characterization of the data in terms of demographics and tables.
[DOCX File , 3055 KB - jmir_v24i5e36431_app1.docx]

References

1. Ruscio AM, Stein DJ, Chiu WT, Kessler RC. The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. Mol Psychiatry 2010 Jan;15(1):53-63 [FREE Full text] [doi: 10.1038/mp.2008.94] [Medline: 18725912]
2. Subramaniam M, Soh P, Vaingankar JA, Picco L, Chong SA. Quality of life in obsessive-compulsive disorder: impact of the disorder and of treatment. CNS Drugs 2013 May;27(5):367-383. [doi: 10.1007/s40263-013-0056-z] [Medline: 23580175]
3. Skapinakis P, Caldwell DM, Hollingworth W, Bryden P, Fineberg NA, Salkovskis P, et al. Pharmacological and psychotherapeutic interventions for management of obsessive-compulsive disorder in adults: a systematic review and network meta-analysis. Lancet Psychiatry 2016 Aug;3(8):730-739 [FREE Full text] [doi: 10.1016/S2215-0366(16)30069-4] [Medline: 27318812]
4. McKay D, Sookman D, Neziroglu F, Wilhelm S, Stein DJ, Kyrios M, et al. Efficacy of cognitive-behavioral therapy for obsessive-compulsive disorder. Psychiatry Res 2015 Feb 28;225(3):236-246. [doi: 10.1016/j.psychres.2014.11.058] [Medline: 25613661]
5. Reid JE, Laws KR, Drummond L, Vismara M, Grancini B, Mpavaenda D, et al. Cognitive behavioural therapy with exposure and response prevention in the treatment of obsessive-compulsive disorder: A systematic review and meta-analysis of randomised controlled trials. Compr Psychiatry 2021 Apr;106:152223 [FREE Full text] [doi: 10.1016/j.comppsych.2021.152223] [Medline: 33618297]
6. Öst L, Havnen A, Hansen B, Kvale G. Cognitive behavioral treatments of obsessive-compulsive disorder. A systematic review and meta-analysis of studies published 1993-2014. Clin Psychol Rev 2015 Aug;40:156-169. [doi: 10.1016/j.cpr.2015.06.003] [Medline: 26117062]
7. Stewart RE, Chambless DL. Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: a meta-analysis of effectiveness studies. J Consult Clin Psychol 2009 Aug;77(4):595-606 [FREE Full text] [doi: 10.1037/a0016032] [Medline: 19634954]
8. Koran LM, Hanna GL, Hollander E, Nestadt G, Simpson HB. Practice Guidelines for the Treatment of Patients With Obsessive-Compulsive Disorder. Psychiatry Online. 2007 Jul. URL: https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/ocd.pdf [accessed 2022-05-10]

XSL•FO
RenderX

9. Obsessive-compulsive disorder and body dysmorphic disorder: treatment. National Institute for Health and Care Excellence. 2005 Nov 29. URL: https://www.nice.org.uk/guidance/cg31 [accessed 2022-05-10]
10. O'Neill J, Feusner J. Cognitive-behavioral therapy for obsessive–compulsive disorder: access to treatment, prediction of long-term outcome with neuroimaging. PRBM 2015 Jul:211-223. [doi: 10.2147/prbm.s75106]
11. Young AS, Klap R, Shoai R, Wells KB. Persistent depression and anxiety in the United States: prevalence and quality of care. Psychiatr Serv 2008 Dec;59(12):1391-1398 [FREE Full text] [doi: 10.1176/appi.ps.59.12.1391] [Medline: 19033165]
12. Foa E, Yadin E, Lichner T. Exposure and Response (Ritual) Prevention for Obsessive Compulsive Disorder: Therapist Guide Internet. New York, US: Oxford University Press; 2012.
13. Wootton BM. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: A meta-analysis. Clin Psychol Rev 2016 Feb;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
14. Lovell K, Cox D, Haddock G, Jones C, Raines D, Garvey R, et al. Telephone administered cognitive behaviour therapy for treatment of obsessive compulsive disorder: randomised controlled non-inferiority trial. BMJ 2006 Oct 28;333(7574):883 [FREE Full text] [doi: 10.1136/bmj.38940.355602.80] [Medline: 16935946]
15. Turner CM, Mataix-Cols D, Lovell K, Krebs G, Lang K, Byford S, et al. Telephone cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled non-inferiority trial. J Am Acad Child Adolesc Psychiatry 2014 Dec;53(12):1298-1307.e2 [FREE Full text] [doi: 10.1016/j.jaac.2014.09.012] [Medline: 25457928]
16. Number of smartphone subscriptions worldwide from 2016 to 2027. Statista. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2022-05-13]
17. Gershkovich M, Middleton R, Hezel DM, Grimaldi S, Renna M, Basaraba C, et al. Integrating Exposure and Response Prevention With a Mobile App to Treat Obsessive-Compulsive Disorder: Feasibility, Acceptability, and Preliminary Effects. Behav Ther 2021 Mar;52(2):394-405. [doi: 10.1016/j.beth.2020.05.001] [Medline: 33622508]
18. Clerkin EM, Teachman BA, Smith-Janik SB. Sudden gains in group cognitive-behavioral therapy for panic disorder. Behav Res Ther 2008 Nov;46(11):1244-1250 [FREE Full text] [doi: 10.1016/j.brat.2008.08.002] [Medline: 18804199]
19. Farrell NR, Ouimet AJ, Rowa K, Soreni N, Swinson RP, McCabe RE. Who gets better when? An investigation of change patterns in group cognitive behavioral therapy for obsessive-compulsive disorder. Journal of Obsessive-Compulsive and Related Disorders 2016 Jul;10:35-41. [doi: 10.1016/j.jocrd.2016.05.003]
20. Steinman SA, Hunter MD, Teachman BA. Do patterns of change during treatment for panic disorder predict future panic symptoms? J Behav Ther Exp Psychiatry 2013 Jun;44(2):150-157 [FREE Full text] [doi: 10.1016/j.jbtep.2012.09.003] [Medline: 23187115]
21. Tolin DF, Gilliam C, Wootton BM, Bowe W, Bragdon LB, Davis E, et al. Psychometric Properties of a Structured Diagnostic Interview for DSM-5 Anxiety, Mood, and Obsessive-Compulsive and Related Disorders. Assessment 2018 Jan;25(1):3-13. [doi: 10.1177/1073191116638410] [Medline: 26988404]
22. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Arlington, VA: American Psychiatric Association; 2013.
23. Abramowitz JS, Deacon BJ, Olatunji BO, Wheaton MG, Berman NC, Losardo D, et al. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychol Assess 2010 Mar;22(1):180-198. [doi: 10.1037/a0018260] [Medline: 20230164]
24. 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] [Medline: 7726811]
25. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of Life Enjoyment and Satisfaction Questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321-326. [Medline: 8290681]
26. Mataix-Cols D, Fernández de la Cruz L, Nordsletten AE, Lenhard F, Isomura K, Simpson HB. Towards an international expert consensus for defining treatment response, remission, recovery and relapse in obsessive-compulsive disorder. World Psychiatry 2016 Feb;15(1):80-81 [FREE Full text] [doi: 10.1002/wps.20299] [Medline: 26833615]
27. Pinto A, Mancebo M, Eisen J, Pagano M, Rasmussen S. The Brown Longitudinal Obsessive Compulsive Study: clinical features and symptoms of the sample at intake. J Clin Psychiatry 2006 May;67(5):703-711 [FREE Full text] [doi: 10.4088/jcp.v67n0503] [Medline: 16841619]
28. Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. Arch Gen Psychiatry 1989 Nov;46(11):1006-1011. [doi: 10.1001/archpsyc.1989.01810110048007] [Medline: 2684084]
29. Feusner JD, Mohideen R, Smith S, Patanam I, Vaitla A, Lam C, et al. Semantic Linkages of Obsessions From an International Obsessive-Compulsive Disorder Mobile App Data Set: Big Data Analytics Study. J Med Internet Res 2021 Jun 21;23(6):e25482 [FREE Full text] [doi: 10.2196/25482] [Medline: 33892466]
30. Asnaani A, Benhamou K, Kaczkurkin AN, Turk-Karan E, Foa EB. Beyond the Constraints of an RCT: Naturalistic Treatment Outcomes for Anxiety-Related Disorders. Behav Ther 2020 May;51(3):434-446 [FREE Full text] [doi: 10.1016/j.beth.2019.07.007] [Medline: 32402259]

XSL•FO
RenderX

Abbreviations

CBT: cognitive-behavioral therapy
DASS: Depression, Anxiety, and Stress Scales
DIAMOND: Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders
DOCS: Dimensional Obsessive-Compulsive Scale
ERP: exposure and response prevention
OCD: obsessive-compulsive disorder

Edited by A Mavragani; submitted 20.01.22; peer-reviewed by M Wu, M Massi, M Mars; comments to author 08.02.22; revised version received 07.03.22; accepted 25.04.22; published 19.05.22.

Please cite as:
Feusner JD, Farrell NR, Kreyling J, McGrath PB, Rhode A, Faneuff T, Lonsway S, Mohideen R, Jurich JE, Trusky L, Smith SM
Online Video Teletherapy Treatment of Obsessive-Compulsive Disorder Using Exposure and Response Prevention: Clinical Outcomes From a Retrospective Longitudinal Observational Study
J Med Internet Res 2022;24(5):e36431
URL: https://www.jmir.org/2022/5/e36431
doi:10.2196/36431
PMID:35587365

©Jamie D Feusner, Nicholas R Farrell, Jeremy Kreyling, Patrick B McGrath, Andreas Rhode, Ted Faneuff, Stephanie Lonsway, Reza Mohideen, John E Jurich, Larry Trusky, Stephen M Smith. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 19.05.2022. 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.

XSL•FO
RenderX

Original Paper

How We Can Reap the Full Benefit of Teleconsultations: Economic Evaluation Combined With a Performance Evaluation Through a Discrete-Event Simulation

Marius Huguet[1], PhD; Marianne Sarazin[1,2,3], MD, PhD; Lionel Perrier[4,5], PhD; Vincent Augusto[1], PhD

[1]Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Centre CIS, F-42023 Saint-Étienne, France

[2]Groupe mutualiste Aesio, F-42100 Saint-Étienne, France

[3]IPLESP- Umrs 1136 Inserm, réseau Sentinelles, F-75012 Paris, France

[4]Univ Lyon, Leon Berard Cancer Centre, GATE UMR 5824, F-69008 Lyon, France

[5]Human and Social Science Department, Centre Léon Bérard, F-69008 Lyon, France

Corresponding Author:
Marius Huguet, PhD
Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Centre CIS
158 Cours Fauriel
Allée 1
F-42023 Saint-Étienne, 42000
France
Phone: 33 0633989861
Email: marius.huguet@gmail.com

Abstract

Background: In recent years, the rapid development of information and communications technology enabled by innovations in videoconferencing solutions and the emergence of connected medical devices has contributed to expanding the scope of application and expediting the development of telemedicine.

Objective: This study evaluates the use of teleconsultations (TCs) for specialist consultations at hospitals in terms of costs, resource consumption, and patient travel time. The key feature of our evaluation framework is the combination of an economic evaluation through a cost analysis and a performance evaluation through a discrete-event simulation (DES) approach.

Methods: Three data sets were used to obtain detailed information on the characteristics of patients, characteristics of patients' residential locations, and usage of telehealth stations. A total of 532 patients who received at least one TC and 18,559 patients who received solely physical consultations (CSs) were included in the initial sample. The TC patients were recruited during a 7-month period (ie, 2020 data) versus 19 months for the CS patients (ie, 2019 and 2020 data). A propensity score matching procedure was applied in the economic evaluation. To identify the best scenarios for reaping the full benefits of TCs, various scenarios depicting different population types and deployment strategies were explored in the DES model. Associated break-even levels were calculated.

Results: The results of the cost evaluation reveal a higher cost for the TC group, mainly induced by higher volumes of (tele)consultations per patient and the substantial initial investment required for TC equipment. On average, the total cost per patient over 298 days of follow-up was €356.37 (US $392) per TC patient and €305.18 (US $336) per CS patient. However, the incremental cost of TCs was not statistically significant: €356.37 – €305.18 = €51.19 or US $392 – US $336 = US $56 (95% CI –35.99 to 114.25; *P*=.18). Sensitivity analysis suggested heterogeneous economic profitability levels within subpopulations and based on the intensity of use of TC solutions. In fact, the DES model results show that TCs could be a cost-saving strategy in some cases, depending on population characteristics, the amortization speed of telehealth equipment, and the locations of telehealth stations.

Conclusions: The use of TCs has the potential to lead to a major organizational change in the health care system in the near future. Nevertheless, TC performance is strongly related to the context and deployment strategy involved.

(J Med Internet Res 2022;24(5):e32002) doi:10.2196/32002

XSL•FO
RenderX

KEYWORDS

telemedicine; telehealth; teleconsultation; discrete-event simulation; economic evaluation; propensity score matching

Introduction

In recent years, the rapid development of information and communications technology (ICT) has contributed to expanding the scope of application and expediting the development of telemedicine [1]. Initially, telemedicine was mainly developed to tackle medical desertification in remote areas and as a useful tool in the context of natural and human-made disasters [2-4]. The primary objective of telemedicine was to enhance patients' access to care by offering the opportunity for patients located in remote areas to gain fast and distant access to medical expertise through the use of ICT [3]. More recently, the COVID-19 crisis has been a turning point in the development of telemedicine and has drastically accelerated its adoption. The use of telemedicine has soared, with 486,369 teleconsultations (TCs) performed in France during the last week of March 2020, against an average of 10,000 TCs per week before the COVID-19 crisis [5].

Telemedicine is a generic term gathering several heterogeneous subcategories, such as TC, telemonitoring, teleassistance, and tele-expertise [6]. Focusing on TCs, one can further distinguish between video TC and TC using a telehealth station [7]. The former refers to a TC involving a patient and a distant physician using a videoconferencing solution. Videoconferencing solutions dedicated to TCs are software programs that include various types of services. The latter refers to a TC involving, on the one hand, a patient accompanied by a medical professional and, on the other hand, a distant physician. A telehealth station is a piece of hardware that includes a videoconferencing solution and a variety of connected medical devices (eg, a stethoscope, a handheld camera, an ultrasound scanner, an electrocardiogram). The connected medical devices are handled by the medical professional who is accompanying the patient, and they allow real-time transmission of data to the distant physician for interpretation. In France, many telehealth stations have been installed in nursing homes or in drug stores. This process has been facilitated by rider 6 to the French agreement for nurses, which recognized accompaniment by nurses as a medical act with an associated fee per TC [8]. Similarly, rider 15 to the French convention of pharmacists sets a flat-rate remuneration based on the annual volume of TCs completed [9].

In the scientific literature, the use of TCs has been intensively studied from a variety of perspectives and in a variety of contexts. Many studies have questioned the relative quality of care of a TC compared with a physical consultation (CS) by conducting randomized controlled trials. A few of these studies have used generic health measures such as quality-adjusted life years [10,11]. Nevertheless, TCs were found to have no impact on such long-term patient outcomes. To increase the likelihood of detecting small changes in quality of care, other studies have relied on various disease-specific measures (eg, diagnostic accuracy, reduction in wound size, blood glucose level) [11-13]. Overall, these studies have tended to demonstrate that the use of TCs is a safe alternative to CSs, providing a noninferior level of quality in a variety of contexts. The question of quality of care has also been investigated from the patient perspective through questionnaires concerning satisfaction and patient-reported outcomes. A recent randomized controlled trial on orthopedic consultations based on the 3-level version of the EuroQol 5-dimensional system (EQ-5D-3L) found no difference in perceived quality of care [14]. Regarding patient satisfaction with TCs, several surveys have been conducted (ie, both disease specific or nondisease specific), indicating a high level of satisfaction [15-17].

Many economic evaluations have also been conducted through cost analysis or cost-effectiveness analysis. Five reviews of the existing literature on economic evaluations addressing studies prior to 2010 failed to reach any reliable conclusion, arguing that economic evaluations of telemedicine were less adherent to methodological standards than evaluations in other fields (eg, featuring a lack of information on the costing methodology, the perspective of the evaluation, and sensitivity analysis) [6,18-21]. In a recent literature review focusing on telemonitoring, the French Authority for Health observed a substantial increase in the methodological quality of evaluations compared with a previous review conducted by the Authority in 2013 [6,22]. More recently, 2 literature reviews addressed studies published during the period 2014-2020. One of these was a scoping review that included 50 economic evaluations of telemedicine [23], which was found to result in cost savings of 53%, 50%, and 32% in cost-minimization, cost-effectiveness, and cost-utility analyses, respectively. These analyses tended to identify increased productivity through a reduction in consultation time; however, such a reduction might be offset by the associated increase in administrative overhead. Furthermore, a reduction in resource consumption is unlikely to result in cost savings under an activity-based payment scheme. The scoping review substantiated several scenarios in which telemedicine could lead to cost savings, such as when medical patient transportation could be avoided. The other literature review was an umbrella review that included 18 systematic reviews on costs or cost-effectiveness analysis [24]. Among the 18 systematic reviews included, 7 concluded that telemedicine was cost saving, 4 concluded that telemedicine was more expensive, and the remaining 7 reviews were unable to reach a conclusion due to heterogeneity in the outcome measures and the poor quality of the cost data. Overall, the heterogeneity in the conclusions among the studies included in these 2 literature reviews may partially be explained by the variety of diseases and contexts in which telemedicine was evaluated. Moreover, the multitude of diseases investigated independently in a disease-specific setting prevented many studies from considering the amortization of telemedicine equipment as the rate of equipment utilization for other diseases was unknown.

In this context, the aim of this study is twofold. First, this study intends to provide an economic evaluation of the use of TCs for specialist consultation at hospitals through a cost analysis. Second, we conduct a performance evaluation based on a discrete-event simulation (DES) with key performance indicators

XSL•FO
RenderX

(KPIs) such as costs, travel time, and resource time consumption.

In fact, several studies have employed simulation and modeling techniques to assess the performance and facilitate the deployment of telemedicine in a variety of contexts. Modeling techniques such as Petri nets have been used to describe health care systems formally with the purpose of performance evaluation. A generic modeling approach to alarm management workflows in health care was proposed by Fanti et al [25] using UML (uniform modeling language) for communication and colored timed Petri nets for simulation. The framework was shown to have high potential capability for describing large and complex health care systems. Dotoli et al [26] proposed a continuous Petri net framework to describe the structure and dynamics of an emergency cardiology department. In the same paper, the model allowed for the generation of an optimization problem and a simulation model. Dotoli et al [27] also used UML activity diagrams and Petri nets to improve the management of hospital departments, and proposed a case study on a pulmonary department. The authors claimed that their base model could be used to design and size any hospital department. Hamana et al [28] used Petri nets to model patient care pathways along with information flows. The authors proposed a performance evaluation approach through Petri net simulation, taking into account degraded modes related to information communication problems.

Regarding simulation, such an approach was used to facilitate the deployment of a telemedicine program in Mexico that consisted of a mobile unit aimed at providing TCs with a distant expert to people in extreme poverty or remote locations [29]. Based on a DES, the authors built a flexible model to calibrate the resources (eg, physicians, mobile units, satellite coverage) to increase the program's utilization rate. Similarly, Qiao et al [30] proposed a DES model to calibrate the resources (eg, TC rooms at hospitals and physicians) to minimize patient waiting times. Considering empirical TC flows from the Henan Telemedicine Center of China, they proposed an optimal sizing of the resources for that hospital. From a theoretical perspective, another study explored the optimal allocation of resources (eg, TC rooms, experts) that would minimize waiting times in the provision of TCs [31]. Based on queuing theory, the findings of the study indicated that the combination of the number of experts and TC rooms does indeed have a decisive impact on the queue length and that the impact of TC rooms is much larger. In a different context, a study investigated the use of TCs as a way for specialists to review patient referrals to remove inappropriate patients from specialist queues [32]. Using a DES approach and data from a rheumatology clinic, the authors found that without TCs, lead times were very long, and the use of TCs as a triage tool was found to be very effective in increasing the performance of the system. Based on a French experiment on telemedicine in geriatrics launched in 2006, a study employed a system dynamics approach through a parametric scenario–based model to compare the performance of a system without telemedicine with 3 alternative scenarios involving the use of telemedicine (eg, TCs only, tele-expertise only, and both TCs and tele-expertise) [33]. Assuming that quality of care with telemedicine and quality of care without telemedicine are comparable, they considered KPIs such as total health care costs, carbon dioxide equivalent emitted, and total medical time available. Their findings favored the tele-expertise scenario for increasing the total medical time available, while the scenario combining tele-expertise and TCs tended to be superior in terms of total costs and environmental aspects.

The uniqueness of our study lies in part in the combination of an economic evaluation of TCs through a cost evaluation and a performance evaluation that uses a DES approach. The cost analysis seeks to determine whether the way in which TCs were deployed at a much larger scale during the early stages of the COVID-19 pandemic (ie, early 2020) was a cost-saving strategy, and the DES approach explores the performance associated with several alternative scenarios for future deployment. The DES approach also substantiates the importance of resources other than cost in the evaluation, such as medical time and administrative time. Our evaluation framework is not disease specific and, instead, considers any eligible specialty. Our study thus contributes to the literature on economic evaluations by investigating in more detail the issue of telemedicine equipment amortization by taking into account all specialties that might benefit from investments in this equipment. Moreover, we compare patients' care pathways over a 298-day period of follow-up in the economic evaluation, which allows us to shed light on changes in the demand for (tele)consultations. Furthermore, to the best of our knowledge, our study is the first to distinguish between TCs using a videoconferencing solution and TCs using a telehealth station. This distinction allows us to investigate various scenarios for the future deployment of TCs based on the relative intensity of use of CSs, video TCs, and TCs using a telehealth station within the eligible population. However, we are able to make this distinction only in the DES approach; in the economic evaluation, video TCs and TCs using a telehealth station are merged into a single group as it was not possible to obtain the relevant information at the individual level in the data. Our approach substantially differs from that of studies investigating the best allocation of resources given a flow of TCs because we evaluate the performance associated with various levels of intensity of TC use (ie, which lead to different TC flows) given a population.

Methods

Positioning of the Problem

In this paper, we propose a generic model of TC use, taking into account 3 alternatives: (1) classical CSs (pathway 1), (2) TCs using a videoconferencing solution (pathway 2), and (3) TCs using a telehealth station (pathway 3). To formally define the corresponding patient pathways, we propose a Petri net model illustrated in Figure 1. Pathways 1, 2, and 3 are represented using a different color: blue, green, and yellow, respectively.

Source transition t_0 models the arrival of consultation requests in the system. Place p_0 models the choice between classical CSs (the patient is not eligible for a TC, transition t_1), TCs using a videoconferencing solution (transition t'_1), and TCs using a

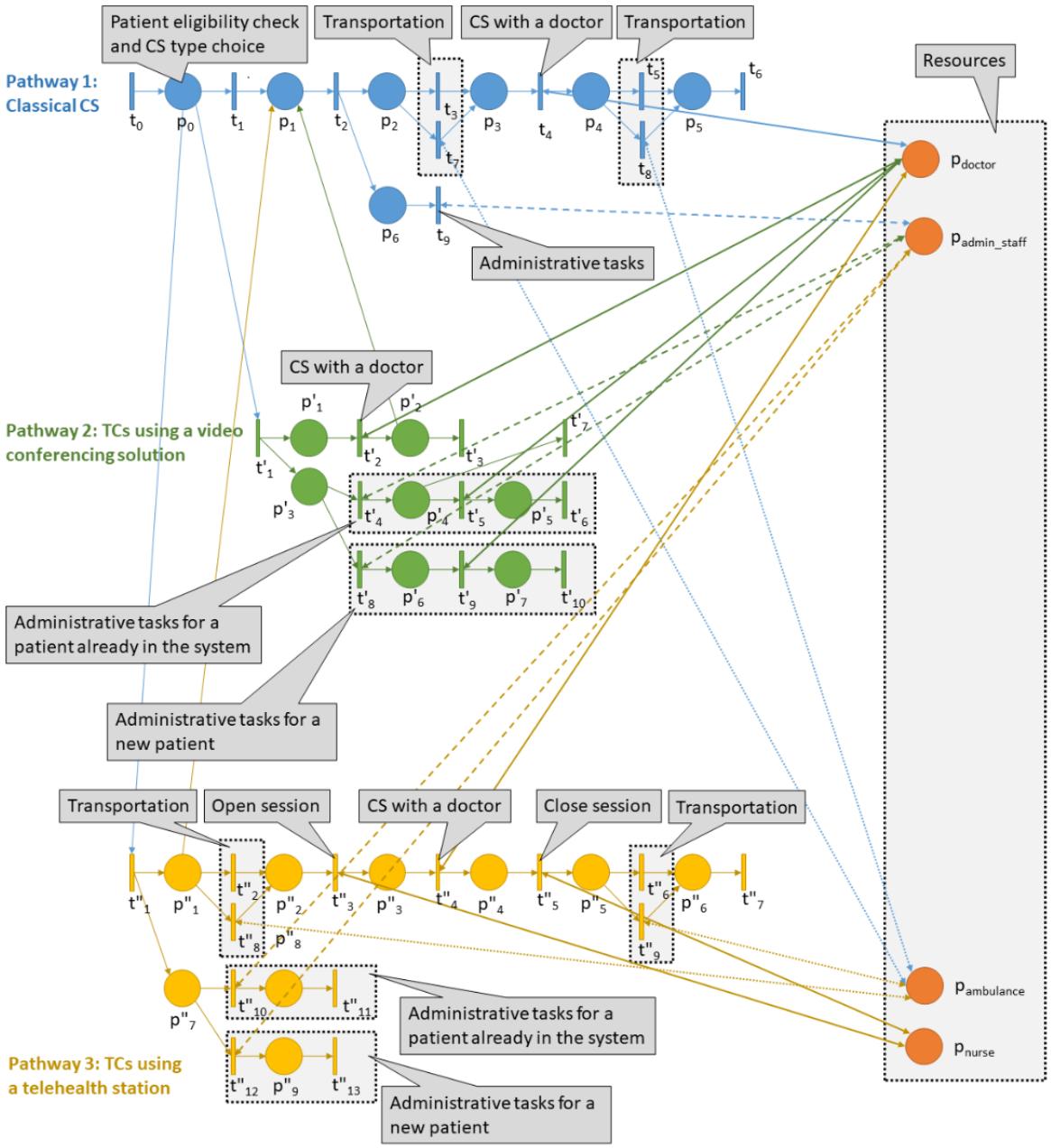
XSL•FO
RenderX

telehealth station (transition t''_1). Places p_{doctor}, p_{admin_staff}, $p_{ambulance}$, and p_{nurse} model the resources taken into account.

In pathway 1, transition t_2 models the activation of 2 parallel activities related to the consultation itself (pathway p_2-t_6) and administrative work (pathway p_6-t_9). The patient goes to the CS using his/her personal vehicle (transitions t_3 and t_5) or an ambulance (transitions t_7 and t_8). Transition t_4 models the CS with a doctor. The patient then exits the system (sink transition t_6). Finally, transition t_9 models administrative tasks executed in the background.

In pathway 2, transition t'_1 models the activation of 2 parallel activities related to the consultation itself (p'_1-t'_3) and administrative work (p'_3-t'_{10}). The video TC is performed with a doctor (transition t'_2). Then, the patient may need a classical CS and be redirected to place p_1 in pathway 1. Otherwise, the patient exits the system (sink transition t'_3). If the patient is new (pathway p'_3-t'_7), the administrative tasks and consultation are longer than the pathway (t'_8-t'_{10}).

Figure 1. Petri net of patient pathways. CS: physical consultation; TC: teleconsultation.



In pathway 3, transition t''_1 models the activation of 2 parallel activities related to the consultation itself (pathway p''_1-t''_7) and administrative work (pathway p''_7-t''_{13}). If the nurse is sick (place p''_1), the patient is redirected to a classical CS. Otherwise, the patient goes to the equipment location using a personal vehicle (transitions t''_2 and t''_6) or an ambulance (transitions t''_8 and t''_9). Then, the nurse sets up the equipment and opens the session (transition t''_3), the doctor performs the consultation (transition t''_4), and the nurse closes the session (transition t''_5). The patient then exits the system (sink transition t''_7). Finally, depending

XSL•FO
RenderX

on whether the patient is new (t''_{10}-t''_{11}) or not (t''_{12}-t''_{13}), administrative tasks with different durations are performed.

Based on this formal model, the problem addressed in this paper consists of proposing control policies to optimize the operations of the global system. Such control policies are strongly connected to decision p_0, that is, how to assign a pathway to a patient taking into account his/her characteristics. Traveling durations and the amount of available resources (doctors, administrative staff, ambulances, and nurses) are also taken into account. The proposed model followed the recommendation of the French Authority for Health to capture organizational changes induced by health technologies, for example, by considering changes in processes (eg, the appointment process, consultation process) induced by the use of TCs [34].

Data

In this study, we used 3 data sets to obtain detailed information on the characteristics of patients, the characteristics of patients' residential locations, and the usage of telehealth stations. First, we identified all patients who had at least one CS or TC with a specialist at the *Clinique Mutualiste de Saint-Etienne* private hospital from January 2019 to September 2020. This clinic specializes in surgeries related to fields such as urology, orthopedics, digestive medicine, gynecology, and treatment of obesity, with 15% of its practice dedicated to cancer. The clinic also has a geriatric department and a dementia center. From January 2019 to March 2020 (the beginning of the COVID-19 pandemic in France), telemedicine was not part of Clinique Mutualiste medical practice. As the pandemic increased in severity, all French hospitals were forced to limit care to patients with COVID-19 or urgent illness. Therefore, all surgical interventions and care for low emergency cases were postponed. To stay in touch with patients and to ensure psychological follow-up, especially if the patients had cancer, obesity problems, or dementia or needed anesthetic advice before a surgery that cannot be postponed, Clinique Mutualiste established a telemedicine program. This program consisted of 2 telemedicine cabins with competent nurses outside the clinic for patients who needed a total examination and special visual software that could be downloaded by patients on their phone or computer. All practitioners working in all medical specialties were taken into account. Other telehealth stations were also introduced in 10 nursing homes around the clinic within the Loire department.

The inclusion criteria were having at least one (tele)consultation during the period of follow-up and that the (tele)consultations were performed before or after a hospital stay. The link with a hospital stay is required to obtain information on patient characteristics from patient records. We relied on a fixed period of follow-up to have the same chance of observing a (tele)consultation for each patient. To maximize the number of TCs included in the sample, we retained the period of follow-up of 165 days before hospitalization and 133 days after hospitalization (see Multimedia Appendix 1 for more details). For each patient, the information recorded included the residential location of the patient, his/her age, gender, ICD-10 (International Statistical Classification of Disease and Related Health Problems, tenth revision) diagnostic codes (see Multimedia Appendix 2 for more details), care unit (eg, ambulatory, gynecology, urology, neurology), and date of discharge as well as the date of each (tele)consultation. We also computed the travel distance and travel time by car between patients' residential municipalities and the private hospital.

Second, we included aggregate information about patients' residential locations from open access data sets managed by the *National Institute of Statistics and Economic Studies* (INSEE). We obtained information about the municipalities, such as the population density and median standard of living (in euros) within each municipality in mainland France.

Finally, we included data on the usage of telehealth stations in the department of Loire from an application programming interface (API) provided by HOPI Medical. The information recorded included the date of each TC performed using a telehealth station and the TC time (minutes).

Ethics Approval

This study was conducted in accordance with the ethical principles for medical research involving human subjects developed in the Declaration of Helsinki by the World Medical Association (WMA). The study received approval in France from the National Ethics Committee (CESREES N°2809078 bis, CNIL N°921041).

Economic Evaluation (Cost Analysis)

We first conducted a cost analysis of the use of TCs versus CSs over 298 days of follow-up (see Multimedia Appendix 1). The analysis was conducted from a health care system perspective. In fact, using a retrospective data set, it was not possible to retrieve data on informal care, which prevented us from considering a collective perspective. Based on the recommendations of the French Authority for Health, costs were not discounted since the time horizon was less than 12 months [35].

The 2 strategies being compared are as follows:

- TC patients: patients having at least one TC during the follow-up period. TCs performed using a videoconferencing solution or a telehealth station were merged in the same group as it was not possible to distinguish between the 2 types of TCs at the individual level in the data.
- CS patients (the control group): patients having solely CSs during the follow-up period.

An important feature of the economic evaluation is that we evaluated the cost differences over patients' care pathways (ie, 298 days of follow-up). Therefore, a TC patient might also have had CSs during his/her care pathway, and these CSs are taken into account in the total cost. By doing so, we evaluated a mixed organization of care in which patients might be treated with both TCs and CSs during their care pathway. The incremental cost between our 2 groups of interest is computed. Based on the recommendations of the French Authority for Health, we considered direct production costs only (ie, costs of the resources required for the production of the interventions evaluated). All costs are valued in euros (2020 data). The total cost for patient *i* is computed as follows based on the cost inputs presented in Table 1:

XSL•FO
RenderX



where $NbConsult_i$ is the number of (tele)consultations (all specialties) during the period of follow-up; $p^{TC\ Station}$ is the probability that a TC was performed using a telehealth station; $NbTC_i$ is the number of TCs (all specialties) during the period of follow-up; $p^{medTravel}$ is the share of medical patient transportation; p^{VSL}, p^{Taxi}, and p^{Amb} are the probability of using a light health vehicle (VSL), a taxi, or an ambulance, respectively, if a patient uses medical transportation; $dist_i$ is the cumulative distance traveled by patient *i* during the period of follow-up; VSL(), TAXI(), and AMB() are the functions returning the cost of transportation by a VSL, a taxi, and an ambulance, respectively, for a specified distance (see Multimedia Appendix 3 for more details); NbTeleStation is the number of telehealth stations; $MonthlyEquipment^{Device}$ includes the cost of a full telehealth station, cost of installation, cost of consumable medical equipment, and cost of medical staff training. To take into account the period of usage and the depreciation of the TC equipment, this cost is converted into a monthly cost over the entire period of usage (eg, 60 months if full depreciation at 5 years is assumed); $MonthlyEquipment^{Wallet}$ includes the cost of a telehealth station (wallet), cost of consumable medical equipment, and cost of medical staff training; MaintenanceCost is the cost of maintenance of a telehealth station (monthly); MedEquipment is the medical equipment per telehealth station; NbMonth is the number of months (time window) of the study; N^{TC} is the number of patients who had at least one TC. The frequency was computed before any matching procedure or censoring procedure (ie, period of follow-up) to take into account the real/observed TC usage during the study period; NbDoctors is the number of doctors trained to use TCs within the hospital; MonthlySoftware is the cost of the videoconferencing solution (per month and per user); and CostHeadphone is the cost of a headphone and camera per user.

XSL•FO
RenderX

Table 1. Main cost inputs and other inputs (in euros [2020 data])[a].

Inputs	Value	Sources
TC[b] station equipment costs		Financial data, private hospital in Saint-Etienne
Telehealth station, full station (per station)	21,686.4	
Telehealth station, wallet (per station)	10,200	
Medical equipment (per station)	360	
Installation cost (per station)	1680	
Training (per team of 4 medical professionals)	1020	
Maintenance (per month per station)	90	
Consumable medical equipment (per year or per 200 TCs)	360	
Video TC equipment costs		Financial data, private hospital in Saint-Etienne
Software and storage server (per month per user)	70	
Camera and headphone (per user)	110	
Patient transportation costs		
Nonmedical patient transportation (cost/kilometer)	0.523	[36]
Medical patient transportation	See Multimedia Appendix 3	Multimedia Appendix 3
(Tele)consultation costs (per consultation)		
Average cost for a specialist (tele)consultation (including average out-of-pocket fees)[a]	35.83	[37]
Tariff for nurse accompaniment during a TC	12	[8]
Other inputs		
Share of medical patient transportation	0.36	Assumption
Number of telehealth stations	12	Data, nursing homes in the department of Loire (Aésio group)
Share of telehealth stations with connected devices	0.5	Data, nursing homes in the department of Loire (Aésio group)
Number of teams of 4 medical professionals per telehealth station	1	Data, nursing homes in the department of Loire (Aésio group)
Depreciation rate of TC equipment	5 years	Assumption
Probability that a TC was performed using a telehealth station ($P^{\text{TC Station}}$)	0.2656	Data, HOPI Medical
Number of doctors using TCs	30	Data, private hospital in Saint-Etienne

[a]Costs were reported in 2017 euros in the report from the French National Health Insurance, and were expressed in 2020 euros using the international Classification of Individual Consumption by Purpose (COICOP 06.2.1.2.1) discount rate. €1 = US $1.1 (2020 data).

[b]TC: teleconsultation.

The cost evaluation for CS patients includes the average cost for a specialist CS (including average out-of-pocket fees) multiplied by the number of CSs observed during the period of follow-up as well as the cost of patient transportation. We computed the average cost for a specialist CS, including average out-of-pocket fees, based on open access data for 2017 taken from the French National Health Insurance report [37]. Since the latest data set available was from 2017, the average cost in 2020 euros was expressed using the COICOP 06.2.1.2.1 discount rate. Nonmedical transportation was valued using the official cost per kilometer according to French legislation [36]. Medical transportation was valued for each transportation mode (eg, VSL, Taxi, Ambulance) according to the conventional tariffs set by the French National Health Insurance [38-40]. Then, the average medical transportation cost was computed as the sum of each transportation cost weighted by its probability of use. The probability of use of each transportation mode was derived from a report by the French Directorate for Research, Studies, Assessment, and Statistics (DREES) based on 2018 health expenses data [41]. Finally, transportation costs were computed as the sum of medical and nonmedical transportation costs weighted by the assumed share of patients using medical transportation. For TC patients, the cost evaluation further included TC equipment costs, the training of medical staff, and the fee for nurse accompaniment in cases where a telehealth station was used. Cost inputs were retrieved from financial data taken from the private hospital in Saint-Etienne (France). To overcome the problem posed by the fact that we could not

XSL•FO
RenderX

distinguish between the 2 types of TCs (ie, video versus telehealth station) in the data, the tariff for nurse accompaniment was weighted with the probability that a TC was performed using a telehealth station. While we do not observe the type of TC (ie, video versus telehealth station) at the individual level, data from HOPI Medical allow us to compute the share of TCs performed using a telehealth station. The investment in TC equipment is assumed to be fully depreciated in 5 years, and we assume in the economic evaluation a constant rate of utilization of TC equipment during that period. Additionally, the cost of TC equipment is spread over the real number of patients who had a TC before any exclusion criteria were applied.

In the comparison of costs between our 2 groups of interest, selection bias is likely to occur, induced by the fact that not all patients are eligible for the use of TCs, depending on their individual or disease characteristics. In the economic evaluation, we need to control for this selection bias because the potential differences in characteristics could introduce bias when evaluating the costs. To that end, we rely on a propensity score matching procedure to select a representative control group [42,43]. The propensity score is the conditional probability that a patient will be part of the treatment group (ie, part of the TC patient group), conditional on observable characteristics. This conditional probability is used as a unidimensional indicator of patient characteristics. In other words, 2 patients with a similar propensity score should have similar characteristics (ie, characteristics involved in the estimation of the propensity score). We determined this probability by fitting the following logit model through maximum likelihood estimation:



where the latent variable $y^* = X'\beta + $; the parameter vector β is obtained through maximum likelihood estimation; and follows a logistic distribution. The matrix of patient characteristics X' includes age, gender, the travel time to the hospital, the median standard of living, population density, the care pathway type, the ICD-10 chapter, and the care unit. The decision regarding which variables to include in the logit model should not be based on their expected predictive power. If the propensity score model was designed as a classifying model and assuming a high accuracy of that model, the matching procedure would fail to balance patient characteristics because the patients in the treatment group would have a probability close to 1 and the patients in the control group would have a probability close to 0. Instead, one should include all variables suspected of inducing selection bias between the 2 groups being compared. We then conducted a 1:1 matching procedure by matching each TC patient to the closest CS patient in terms of the propensity score.

To take into account the uncertainty surrounding the point estimation of the incremental cost, we conducted a probabilistic sensitivity analysis. We used a nonparametric bootstrap procedure with 1000 replications [44]. Nonparametric bootstrapping is a resampling procedure where each bootstrap sample is generated by a random sample with replacement from the initial data set. This method is widely used in statistics to obtain the distribution of a point estimate. We also conducted a deterministic sensitivity analysis to explore the sensitivity of the mean difference in cost to the variations in several cost inputs through a tornado diagram.

Performance Evaluation (DES)

We implemented the Petri net model presented in Figure 1 using a DES approach. The population was generated by randomly selecting each agent with replacement from the 2020 clinical database described in the "Data" section. The model was simulated over a 5-year period (253 working days per year) to cope with the assumed amortization time of TC equipment. The eligibility criteria for TC patients were based on agents' attributes (eg, ICD-10 chapters, care units), excluding from the eligible population attributes never observed for TC patients in the data. Finally, conditional on the eligible population, agents were dispatched among the 3 consultation types based on probabilities. Considering that patients older than 80 years old are less likely to be eligible for video TCs, we assumed that for this subpopulation, the probability P^{visio} is upper bounded at 0.3. Thus, in scenarios in which P^{visio} exceeds 0.3, the residual probability $P^{visio} - 0.3$ is attributed equivalently to P^{CS} and $P^{TC\ Station}$. Table 2 provides a full description of the calibration of the parameters.

In the simulation, we assumed a maximum capacity of resources (medical transportation, doctor, nurse, and administrative staff times) and tracked their level of use. For telehealth stations, however, we specified a limited capacity. We assumed that each station could handle a maximum of 7 TCs per day. When the daily workflow exceeds the maximum capacity, the agent is redirected to a CS.

This study is not restricted to a single specialty or disease, and it aims to evaluate the use of TCs from an organizational perspective. Therefore, in accordance with the existing literature (see the "Introduction" section), we assume that the clinical effectiveness of TCs and CSs is comparable among patients eligible for TC [10-14]. This assumption does not negate the fact that not all patients are eligible for the use of TCs, depending on individual or disease characteristics. Thus, we consider the KPIs in Textbox 1 in the performance evaluation of the system.

We conducted 3 experiments and an extra validation experiment (ie, base scenario), which are summarized in Table 3. In experiment 1, we explored the performance of the model associated with all combinations of probabilities P^{CS}, P^{visio}, and $P^{TC\ Station}$. These probabilities reflect the intensity of use of each type of consultation, and thus, they allow us to evaluate the performance associated with various degrees of deployment of each TC type. As the use of video TCs and TCs using telehealth stations is subject to substantial initial investments (ie, fixed costs), it is a particularly insightful factor in examining variation in performance based on the volume (ie, probabilities) of each type of TC. This experiment will, for example, highlight the minimum volume of video TCs and TCs using a telehealth station required to amortize the videoconferencing system and telehealth stations, respectively.

XSL•FO
RenderX

Table 2. Parameters of the simulation model.

Parameters	Value	Source
Arrival rate (per day)	Poisson(89.953782) in 2020	Clinical data
Probability that a new patient first had a TC[a] (to be multiplied by 1/number of TCs per patient to convert at the consult level)	0.2368	Clinical data (number of first TC/number of TC patients)
Probability of a first TC (not previously registered by the doctor)	0.8512	Clinical data (1/number of TC per patient)
Probability of having a CS[b] (P^{CS})[c]	0.9507	Clinical data
Probability of having a video TC (P^{visio})[d]	0.0362	Clinical data
Probability of having a TC using a telehealth station ($P^{TC\ Station}$)[e]	0.0131	Data, HOPI Medical
CSs		
Travel time (1 way)	Individual agent travel time	Clinical data
Appointment scheduling time	Triangular(1, 2, 1.5)	Secretary staff
CS time (doctor)	Triangular(16.9, 29.1, 20)	[45]
Video TCs		
Consultation time (doctor)	Triangular(8, 9.47, 13.78)	Consult time reduced by 52.6% for otorhinolaryngology [15]
Appointment scheduling time (admin)	Triangular(1, 2, 1.5)	Secretary staff, private hospital in Saint-Etienne
Appointment scheduling time for new patients (admin)	Triangular(8, 10, 9)	Secretary staff, private hospital in Saint-Etienne
Registration time for a first TC (doctor)	Triangular(1, 2, 1.5)	Secretary staff, private hospital in Saint-Etienne
Probability of having a CS within 7 days after a TC	0.0672	Clinical data
TCs using a telehealth station		
Travel time (1 way)	0	Assumption
Preparation time (nurse)	Triangular(8, 12, 10)	Protocol private hospital
Consultation time (doctor)	Lognormal(2.1888426, 0.57548749, 3.0333333)	Data HOPI Medical
Closing time (nurse)	Triangular(3, 7, 5)	Protocol private hospital
Probability of canceling the TC due to the sick leave of a nurse	0.05/number of teams per telehealth station	Assumption
Appointment scheduling time (admin)	Triangular(2.5, 4.5, 3.5)	Department of Anesthesia of a private hospital in Saint-Etienne
Appointment scheduling time for new patients (admin)	Triangular(10, 15, 12.5)	

[a]TC: teleconsultation.

[b]CS: physical consultation.

[c]P^{CS}: CS probability.

[d]P^{visio}: video TC probability.

[e]$P^{TC\ Station}$: TC using a telehealth station probability.

XSL•FO
RenderX

Textbox 1. Key performance indicators considered in the performance evaluation of the system.

Cost key performance indicator

When an agent exits the system, we compute the total cost based on the agent's characteristics (eg, distance traveled) and care pathway based on the same formula used in the economic evaluation. We then sum each total cost over all agents in a simulation run to obtain the overall total cost of the system. Similarly, we compute the average cost associated with each consultation type to obtain the cost per physical consultation (CS), cost per video teleconsultation (TC), and cost per TC using a telehealth station.

Resource usage key performance indicator

Includes doctor, administrative staff, and nurse times (minutes). When an agent consumes a resource, we keep track of the level of consumption by taking a random draw from the distributions in Table 2.

Transfer key performance indicator

The number of patients transferred to a CS after a video TC or after a TC using a telehealth station due to the unavailability of the station or the sick leave of a nurse.

Volume of (tele)consultation key performance indicator

The number of CSs, video TCs, and TCs using a telehealth station completed.

Travel time key performance indicator

The total travel time and travel time avoided (minutes). We compute the travel time avoided for each agent engaging in a TC (ie, either using a video TC or via a telehealth station) as the distance he/she would have traveled to the clinic based on his/her characteristics.

Table 3. Descriptions of the experiments.

Description	Base scenario	Experiment 1	Experiment 2	Experiment 3
Simulation features				
Replication count	20	20	20	20
Replication length (minutes)	1,821,600	1,821,600	1,821,600	1,821,600
Parameters variation				
P^{CSa}	0.9507	Combination summing to one[b]	0 (0.01) 1	Combination summing to one[b]
P^{visioc}	0.0362	Combination summing to one[b]	0	Combination summing to one[b]
$P^{TC\ Stationd}$	0.0131	Combination summing to one[b]	$1 - P^{CS}$	Combination summing to one[b]
Share medical transport	0.36	0.36	1	0.36
Number of telehealth stations	12	12	12	12
Number of teams of nurses per station	1	1	1	1
Population variation				
Poisson daily arrival rate	89.95	89.95	20.64	89.95
Age (years)	No restriction	No restriction	Age >80	No restriction
Distance (km)	No restriction	No restriction	13.06	Distance<20; 20≤Distance<50; Distance≥50

[a]P^{CS}: physical consultation probability.
[b]P^{CS}, P^{visio}, $P^{TC\ Station}$ as long as $P^{CS} + P^{visio} + P^{TC\ Station} = 1$.
[c]P^{visio}: video teleconsultation probability.
[d]$P^{TC\ Station}$: teleconsultation using a telehealth station probability.

The aims of experiments 2 and 3 are similar to those of experiment 1, but these experiments address specific subpopulations. Experiment 2 focused on a population composed of elderly people (ie, age >80 years) in nursing homes. This subpopulation is assumed to not be eligible for video TCs and to be medically transported whenever a CS is needed. Based on the observed distances in our data, we assume that nursing homes are located 13.06 km from the clinic. The primary aim of this experiment was to determine whether the use of telehealth stations is more or less efficient for this subpopulation, considering that these individuals are on average located closer to the clinic but cannot use personal transportation. Experiment 3 split the population into 3 groups: patients living within 20 km, between 20 km and 50 km, and more than 50 km from the clinic, and it replicated experiment 1 for each subpopulation. The objective of this experiment is to shed light on the performance of video TCs or TCs using telehealth stations depending on whether the clinic is attractive in an urban, peri-urban, or remote area and to determine the associated break-even points.

XSL•FO
RenderX

Finally, we conducted an extra validation experiment by simulating the base scenario. In this scenario, the parameters are calibrated to reflect the behaviors observed in the data (eg, the intensity of use of each TC type). This validation experiment serves as a robustness check by comparing the observed distribution of the population characteristics with the simulated distributions.

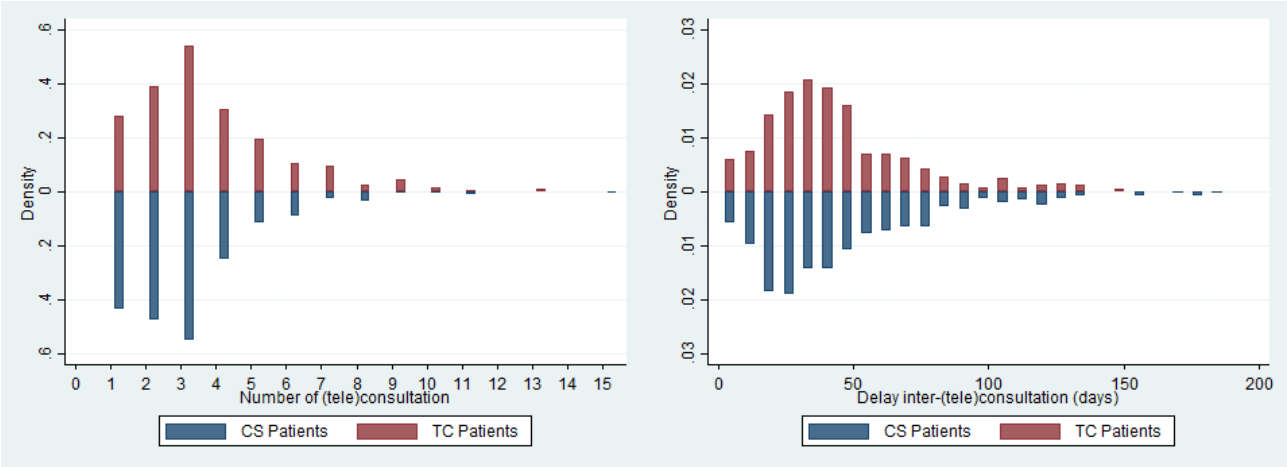
Results

Economic Evaluation (Cost Analysis)

We initially observed 532 patients who received at least one TC (ie, TC patient group) and 18,559 CS patients. The TC patients were recruited during a 7-month period (ie, 2020 data) versus 19 months for the CS patients (ie, 2019 and 2020 data). When censoring the data to have the same period of follow-up for each patient in the sample, the sample size fell to 424 TC patients and 13,202 CS patients. Ultimately, the matched sample used in the economic evaluation included 404 patients in each group. However, we used the initial (full) data set to calibrate the parameters in the DES model.

The number of CSs or TCs is likely to have a substantial impact on the total cost per patient. Figure 2 depicts the density of the number of (tele)consultations and the time between consultations per patient for our 2 groups of interest in the matched sample. The average number of (tele)consultations over the period of follow-up per patient was 2.94 for CS patients and 3.53 for TC patients. Interestingly, an unpaired 2-sample *t* test testing the equality of means indicated a significantly higher number of (tele)consultations for TC patients than for CS patients ($P<.001$). Similarly, the average time between consultations for CS patients was 46.5 days, compared with 43.3 days for TC patients. However, an unpaired 2-sample *t* test indicated no significant difference in means ($P=.17$).

Figure 2. Number of (tele)consultations and time between consultations per patient (matched sample). CS: physical consultation; TC: teleconsultation.



As anticipated, we did indeed observe wide differences in individual or disease characteristics between TC and CS patients, which tended to confirm the presence of selection bias (Table 4). For example, Table 4 shows that 49.1% (208/424) of TC patients engaged in (tele)consultations before and after hospitalization, compared with 36.06% (4761/13,202) of CS patients. This distortion in the consultation care pathway might bias the results if left as it stands because it would artificially increase the relative number of (tele)consultations received by TC patients compared with CS patients. The same issue is likely to occur for the wide differences observed between the 2 groups in ICD-10 chapters and care units.

Table 4 shows a good quality of the propensity score matching procedure in balancing the characteristics between our 2 groups of interest. Indeed, while we initially observed wide differences, after matching by the propensity score, there were no longer any differences in ICD-10 chapters, care units, or CS care pathways between TC and CS patients. The matching procedure inherently balanced the sample size of the 2 groups.

A description of the average total cost computed in the matched sample for TC patients and CS patients is provided in Table 5. On average, the total cost per patient over the period of follow-up was €356.37 (US $392) per TC patient and €305.18 (US $336) per CS patient. Interestingly, regarding the composition of the total cost, the cost associated with the tariff of (tele)consultations is higher for TC patients (ie, by €25 [US $28]) than for CS patients. This result confirms the finding in Figure 2, which shows a significantly higher number of (tele)consultations for TC patients ($P<.001$). By contrast, the average cost of transportation per patient is lower for TC patients (ie, by €72.5 [US $80]) than for CS patients. The cost of transportation for TC patients is greater than 0 because 85.4% of TC patients (362/424) had both CSs and TCs during the period of follow-up. Finally, based on the observed TC utilization rate in the study period, the total cost per TC patient includes an additional cost of €98.76 (US $109) for the TC equipment.

XSL•FO
RenderX

Table 4. Patient characteristics before and after matching by the propensity score.

Patient characteristics	Before matching by the PS[a]			After 1:1 matching by the PS		
	TC[b] patients (n=424)	CS[c] patients (n=13,202)	*P* value[d]	TC patients (n=404)	CS patients (n=404)	*P* value[d]
Age	61.50	61.48	.98	60.93	60.26	.56
Female, n (%)	216 (50.94)	7119 (53.92)	.23	207 (51.24)	192 (47.52)	.29
Male, n (%)	208 (49.06)	6083 (46.08)	.23	197 (48.76)	212 (52.48)	.29
Travel time to hospital (minutes)	24.25	23.21	.38	24.12	25.32	.56
Median standard of living (€[e])	19,765	19,873	.29	19,779	19,921	.33
Population density	9.49	9.62	.81	9.45	10.49	.35
Care pathway consultation, n (%)						
After hospitalization	89 (20.99)	3363 (25.47)	.04	70 (17.32)	67 (16.58)	.78
Before hospitalization	127 (29.95)	5078 (38.46)	<.01	127 (31.44)	121 (29.95)	.65
Both	208 (49.06)	4761 (36.06)	<.01	207 (51.24)	216 (53.47)	.53
ICD-10[f] chapter, n (%)s						
Certain infectious and parasitic diseases	0 (0)	51 (0.39)	.21	0 (0)	0 (0)	—[g]
Neoplasms	100 (23.58)	1679 (12.72)	<.01	95 (23.51)	91 (22.52)	.74
Diseases of the blood and blood-forming organs	0 (0)	15 (0.11)	.51	0 (0)	0 (0)	—
Endocrine, nutritional, and metabolic diseases	15 (3.54)	255 (1.93)	.03	14 (3.47)	17 (4.21)	.58
Mental disorders	0 (0)	1 (0.01)	.81	0 (0)	0 (0)	—
Diseases of the nervous system	1 (0.24)	236 (1.79)	.02	1 (0.25)	2 (0.50)	.56
Diseases of the eye	0 (0)	774 (5.86)	<.01	0 (0)	0 (0)	—
Diseases of the circulatory system	92 (21.70)	1003 (7.60)	<.01	88 (21.78)	90 (22.28)	.87
Diseases of the respiratory system	0 (0)	66 (0.50)	.15	0 (0)	0 (0)	—
Diseases of the digestive system	76 (17.92)	2676 (20.27)	.23	72 (17.82)	79 (19.55)	.53
Diseases of the skin and subcutaneous tissue	1 (0.24)	157 (1.19)	.08	1 (0.25)	1 (0.25)	.99
Diseases of the musculoskeletal system	45 (10.61)	2524 (19.12)	<.01	43 (10.64)	55 (13.61)	.20
Diseases of the genitourinary system	55 (12.97)	1106 (8.38)	.001	52 (12.87)	46 (11.39)	.52
Pregnancy, childbirth, and the puerperium	0 (0)	1 (0.01)	.86	0 (0)	0 (0)	—
Congenital malformations, deformation	2 (0.47)	16 (0.12)	.04	2 (0.50)	1 (0.25)	.56
Symptoms not classified elsewhere	10 (2.36)	412 (3.12)	.46	10 (2.48)	7 (1.73)	.46
Injury, poisoning	1 (0.24)	286 (2.17)	.008	1 (0.25)	0 (0)	.32
Factors influencing health status	26 (6.13)	1944 (14.73)	<.01	25 (6.19)	15 (3.71)	.11
Care unit, n (%)						
Ambulatory emergency	0 (0)	339 (2.57)	.001	0 (0)	0 (0)	—
Ambulatory	143 (33.73)	3427 (25.96)	<.01	143 (35.40)	130 (32.18)	.33
Short stay	55 (12.97)	2397 (18.16)	.006	55 (13.61)	58 (14.36)	.76
Anesthesia	1 (0.24)	1 (0.01)	<.01	0 (0)	0 (0)	—
Other	1 (0.24)	0 (0)	<.01	0 (0)	0 (0)	—
Digestive system	56 (13.21)	1539 (11.66)	.33	48 (11.88)	49 (12.13)	.91
Sleep assessment	2 (0.47)	974 (7.38)	<.01	2 (0.50)	1 (0.25)	.56
Gastroenterology	1 (0.24)	110 (0.83)	.18	1 (0.25)	2 (0.50)	.56
Gynecology	23 (5.42)	566 (4.29)	.26	21 (5.20)	14 (3.47)	.23

XSL•FO
RenderX

Patient characteristics	Before matching by the PS[a]			After 1:1 matching by the PS		
	TC[b] patients (n=424)	CS[c] patients (n=13,202)	*P* value[d]	TC patients (n=404)	CS patients (n=404)	*P* value[d]
Medicine	4 (0.94)	36 (0.27)	.01	3 (0.74)	2 (0.50)	.65
Neurology	7 (1.65)	117 (0.89)	.10	6 (1.50)	5 (1.24)	.76
Ophthalmology	0 (0)	22 (0.17)	.39	0 (0)	0 (0)	—
Orthopedics	19 (4.48)	1741 (13.19)	<.01	19 (4.70)	31 (7.67)	.08
Radiology	0 (0)	1 (0.01)	.86	0 (0)	0 (0)	—
Resuscitation	1 (0.24)	193 (1.46)	.04	0 (0)	0 (0)	—
Monitoring unit	1 (0.24)	384 (2.91)	.001	1 (0.25)	1 (0.25)	.99
Urology	66 (15.57)	712 (5.39)	<.01	61 (15.10)	67 (16.58)	.56
Vascular disease	44 (10.38)	643 (4.87)	<.01	44 (10.89)	44 (10.89)	.99

[a]PS: propensity score.

[b]TC patients: patients having at least one teleconsultation.

[c]CS patients: patients having only physical consultations.

[d]Student *t* test of the difference between TC and CS patients (*P* value); see Multimedia Appendix 2 for a description of the ICD-10 chapters.

[e]€1 = US $1.1 (2020 data).

[f]ICD-10: International Statistical Classification of Disease and Related Health Problems, tenth revision.

[g]No observations.

Table 5. Composition of the total cost (matched sample, in euros [2020 data]).

Composition	TC[a] patients			CS[b] patients		
	Mean (SD)	Range	Share (%)[c]	Mean (SD)	Range	Share (%)[c]
(Tele)consultation cost	130.25 (75.52)	39.02-478.54	36.76	105.27 (65.9)	35.83-537.45	45.56
Medical transportation (1)	87.91 (99.24)	0-1121.65	19.41	132.33 (410.13)	20.44-6964.79	39.47
Nonmedical transportation (2)	39.45 (65.49)	0-831.13	7.89	67.57 (311.37)	2.05-5272.48	14.97
(1)+(2) Total transportation	127.36 (162.85)	0-1952.78	27.3	199.91 (721.12)	22.49-12237.27	54.44
TC equipment	98.76 (0)	N/A[d]	35.94	N/A	N/A	N/A
Total cost per patient[e]	356.37 (213.58)	137.78-2233.88	100	305.18 (753.53)	58.32-12774.72	100

[a]TC: teleconsultation.

[b]CS: physical consultation.

[c]Share: percentage of the total cost.

[d]N/A: not applicable.

[e]Details about the computation of the total cost per patient are provided in the "Methods" section.

The incremental cost of TCs can thus be computed as follows:



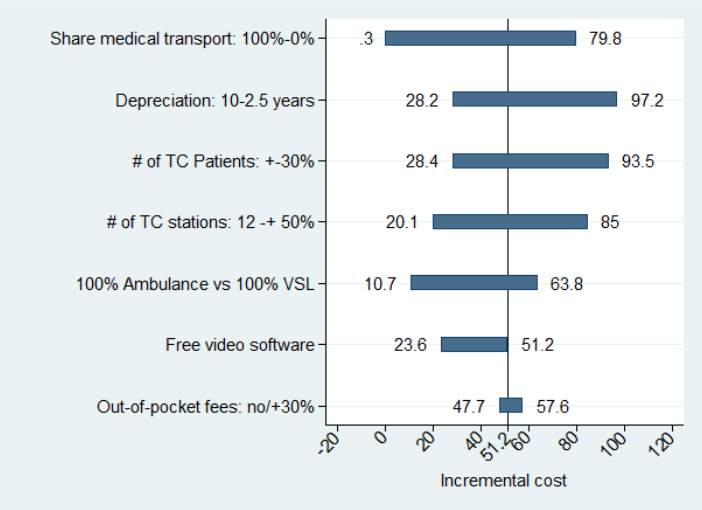
where  are the average of the total costs per TC patient and CS patient, respectively. Overall, the use of TC increased the total cost per patient by €51.19 (US $56) over the period of follow-up.

The probabilistic sensitivity analysis (ie, nonparametric bootstrapping with 1000 replications) allowed us to compute the 95% CI of the point estimation of the incremental cost. The bootstrap procedure captures a strong uncertainty surrounding the incremental cost estimation, which leads to a wide 95% CI.

The results of the deterministic sensitivity analysis are presented as a tornado diagram in Figure 3. The share of patients with medical transportation is found to have the largest effect on the incremental cost. Indeed, the use of TCs would be comparable in terms of total cost if 100% of the patients were assumed to be medically transported. Similarly, assuming a 10-year time to full depreciation of TC equipment, the incremental cost would fall to €28 (US $31) per patient. In the tornado diagram, we included a scenario in which the cost of the videoconferencing software would be free to reflect the economic model of some companies. Indeed, to increase TC station sales, some companies were observed to offer the videoconferencing solution for free as a loss leader product. Under this scenario, the incremental cost would decrease on average to €23.6 (US $26).

XSL•FO
RenderX

Figure 3. Tornado diagram. TC: teleconsultation; VSL: light health vehicle.



Performance Evaluation (DES)

Figure 4 displays the 30 least expensive and the 15 most expensive combinations (P^{CS}, P^{visio}, and $P^{TC\ Station}$) in experiment 1 and indicates their total cost (top) and resource consumption (bottom).

Recall that the triplet of probabilities reflects the intensity of use of each type of consultation. For example, (.3,.5) means that P^{CS}=0.3, P^{visio}=0.5, and $P^{TC\ Station}$=1 – 0.3 – 0.5=0.2. Using this triplet of probabilities, we simulate an organization of care in which on average, among the population eligible for TC, 30% of consultations are CSs, 50% are video TCs, and 20% are TCs using a telehealth station. Thus, by taking all combinations of this triplet, we simulate a multitude of organizations of care that combine the use of CS, video TC, and TC using a telehealth station (see the "Methods" section for more details). The 14 most expensive scenarios depict situations in which there was an investment in telehealth stations ($P^{TC\ Station}$>0) or a videoconferencing solution (ie, P^{visio}>0), but in which the intensity of use of TCs remained very low (ie, <1% of the eligible population). Interestingly, the scenario reflecting the intensity of use of each type of consultation observed in the data (P^{CS}=0.9507, P^{visio}0.0362, $P^{TC\ Station}$=0.0131) is found to be a more expensive organization of care than a scenario without any TCs (P^{CS}=1). Comparing the 14 most expensive scenarios and the scenario without any TCs, one can see that the total cost is substantially lowered in the organization of care with a high intensity of use of video TCs or TCs using telehealth stations. The least expensive scenario (P^{visio}=1) is obtained when all eligible patients are treated through a video TC. Even when P^{CS} is set to 0, there is an incompressible volume of CSs because the probabilities (P^{CS}, P^{visio}, $P^{TC\ Station}$) are conditional on the eligible population.

Regarding this aspect, Figure 5 shows that the incompressible volume of CSs is higher when relying solely on video TCs compared with scenarios mixing the use of video TCs and TCs using telehealth stations. The reason is that elderly people are assumed to be less eligible for video TCs, with a maximum probability P^{visio}=0.3, as well as because of a positive probability of transfer to a CS at the end of a video TC. Thus, the second least expensive scenario (P^{CS}=0, P^{visio}=0.9, $P^{TC\ Station}$=0.1) expands the eligible population and reduces the number of transfers by treating a fraction of the population with TC stations. Additionally, Figure 5 shows that the total number of (tele)consultations completed is slightly higher in scenarios with a higher intensity of video TCs. The reason is that some patients are redirected to a CS after their video TC and due to the lower eligibility of elderly people.

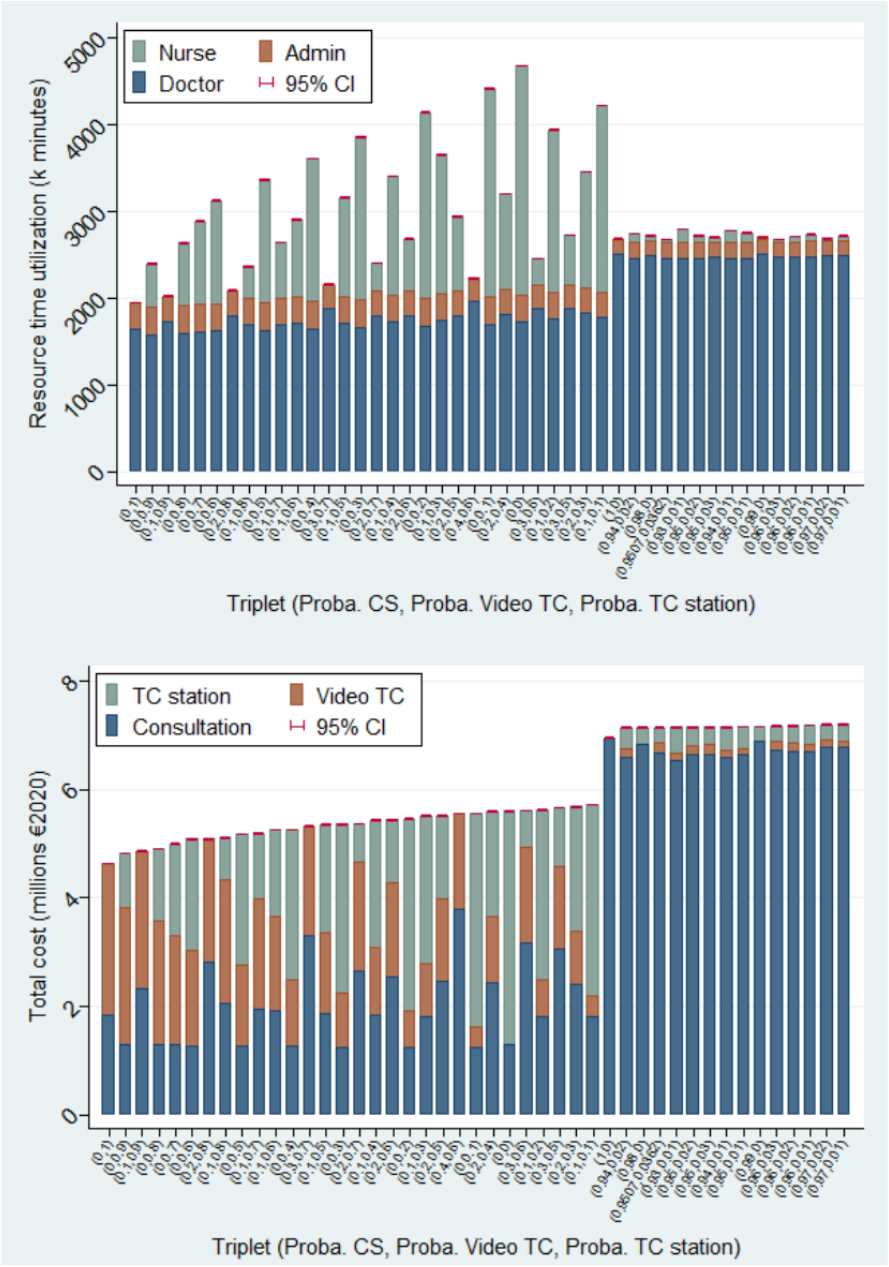
However, the sorting of the scenarios based on their total cost has no relationship with their sorting based on the resource time consumption because such consumption is excluded from the total cost computation from a health care sector perspective (Figure 4, bottom). Our results indicate that the use of TCs is time saving for doctors but is time consuming for administrative staff compared with an organization without any TCs. Overall, we found that the use of a videoconferencing solution is time saving because the decrease in doctor time more than compensates for the increase in administrative time. However, scenarios involving the use of telehealth stations also consume nurse time. Taking into account this extra resource, these scenarios tend to be time consuming in terms of total resources and time due to the substantial consumption of nurse time when $P^{TC\ Station}$>0.2, which largely compensates for the associated decrease in doctor time. Ultimately, the scenario relying solely on video TCs (P^{visio}=1) minimizes the total cost and total resource time consumption while maximizing the total number of (tele)consultations completed for a given population. Nevertheless, relying solely on video TCs does not strictly dominate a combination of video TCs and TCs using telehealth stations because it does not minimize the total travel time KPI (Figure 6). Indeed, the use of TCs (ie, either video TCs or TCs using telehealth stations) naturally leads to substantially lower total travel times compared with CSs. Nevertheless, in this respect, TCs using telehealth stations are even more effective than video TCs. The reason is that a positive share of patients is transferred for CSs after an initial video TC (ie, increased volume) and because of the lower eligibility of elderly people

XSL•FO
RenderX

for video TCs, which expands the incompressible volume of CSs.

Figure 7 depicts the average cost per (tele)consultation type for a given volume of (tele)consultations completed. The average cost per CS is €60.91 (US $67; range 60.01-60.98) and is broadly constant as there is no fixed investment cost for this consultation type. It only varies with the cost of patient transportation depending on the distances traveled to the hospital by patients in the generated population. Thus, CSs are subject to constant returns to scale. By contrast, the average cost per video TC and per TC using a telehealth station drastically decreases over the number of TCs of each type completed. These inverse function shapes (ie, convex shapes) reflect the initial investment (ie, fixed cost) in a videoconferencing solution or in telehealth stations, which are more or less amortized depending on the number of TCs completed. This implies that the total cost functions for video TCs and TCs using telehealth stations are homogenously of degrees lower than 1. In other words, there are substantial scale economies when investing in a videoconferencing solution or a telehealth station. Thus, the minimum volume required to amortize the videoconferencing solution and telehealth station (ie, with 12 stations) investments is 2969 video TCs and 13,604 TCs using a telehealth station, respectively. At these break-even points, the average cost per CS equals the cost per video TC and per TC using a telehealth station. In other words, the videoconferencing solution is amortized if at least 2.3 video TCs per day are completed on average at the hospital (Table 6). Similarly, a telehealth station is amortized if at least 0.9 TCs per station per day are completed.

Figure 4. Bar plot of total cost (upper) and resource time utilization (lower) (experiment 1). CS: physical consultation; TC: teleconsultation.



XSL•FO
RenderX

Figure 5. Bar plot of the volume of (tele)consultations for the scenarios used for experiment 1 (same sorting order as Figure 4). CS: physical consultation; TC: teleconsultation.

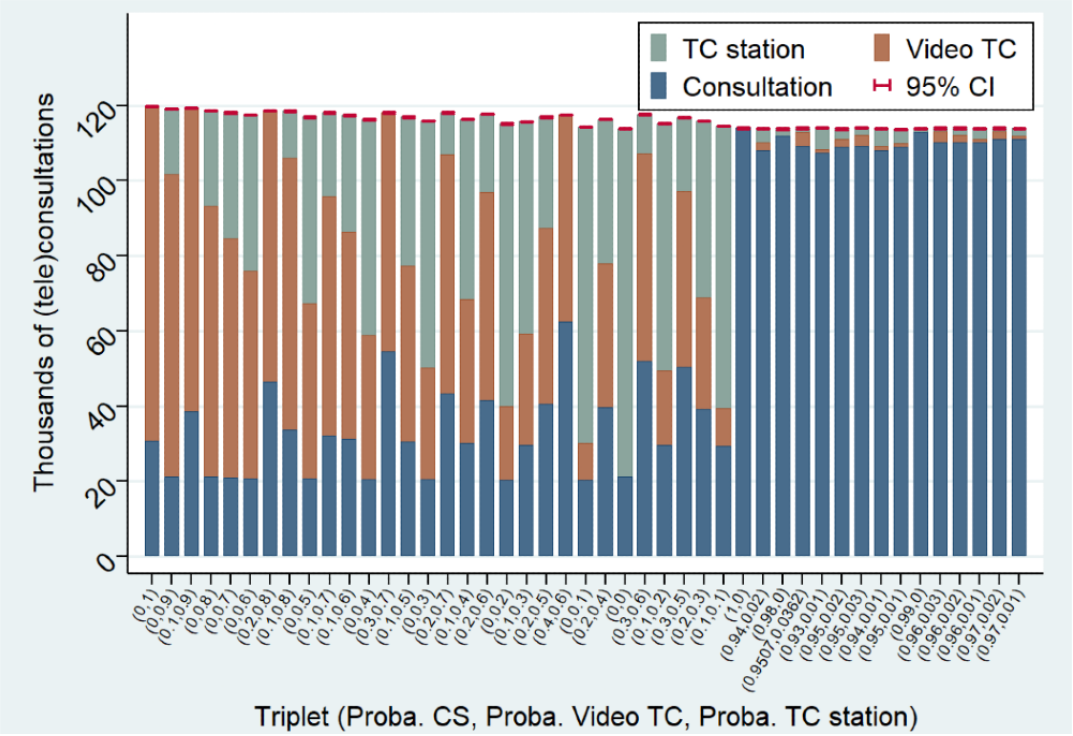
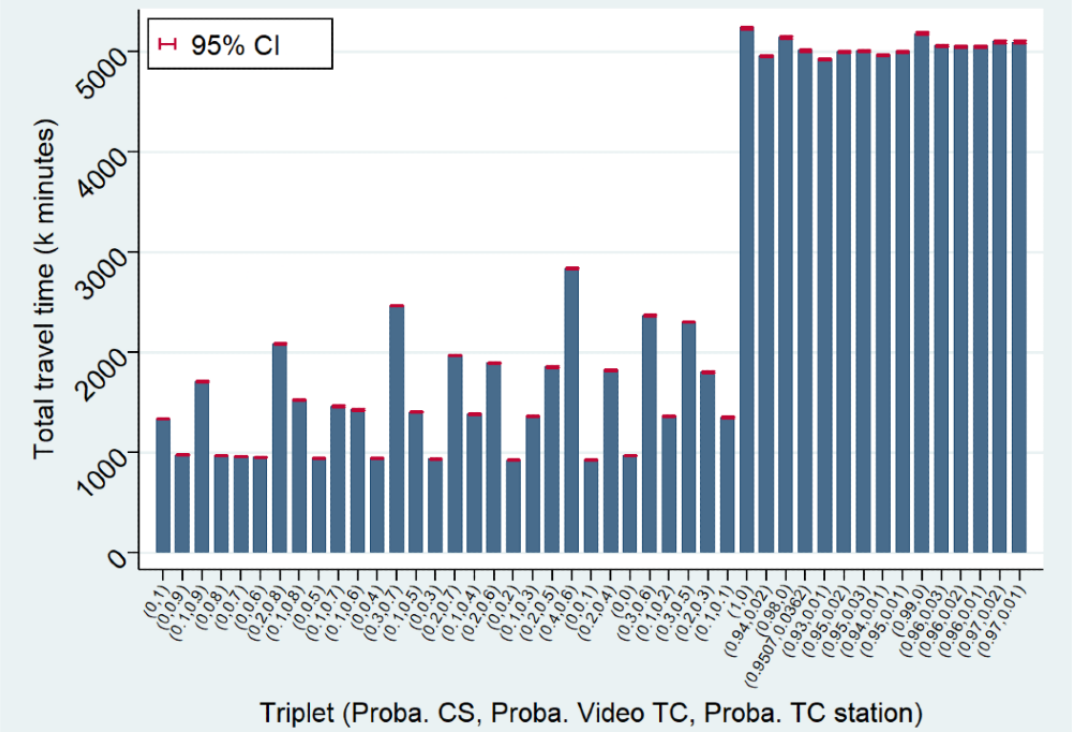


Figure 6. Bar plot of the total travel time (experiment 1). CS: physical consultation; TC: teleconsultation.



XSL•FO
RenderX

Figure 7. Average total cost (cost per (tele)consultation) for the general population (experiment 1). TC: teleconsultation.

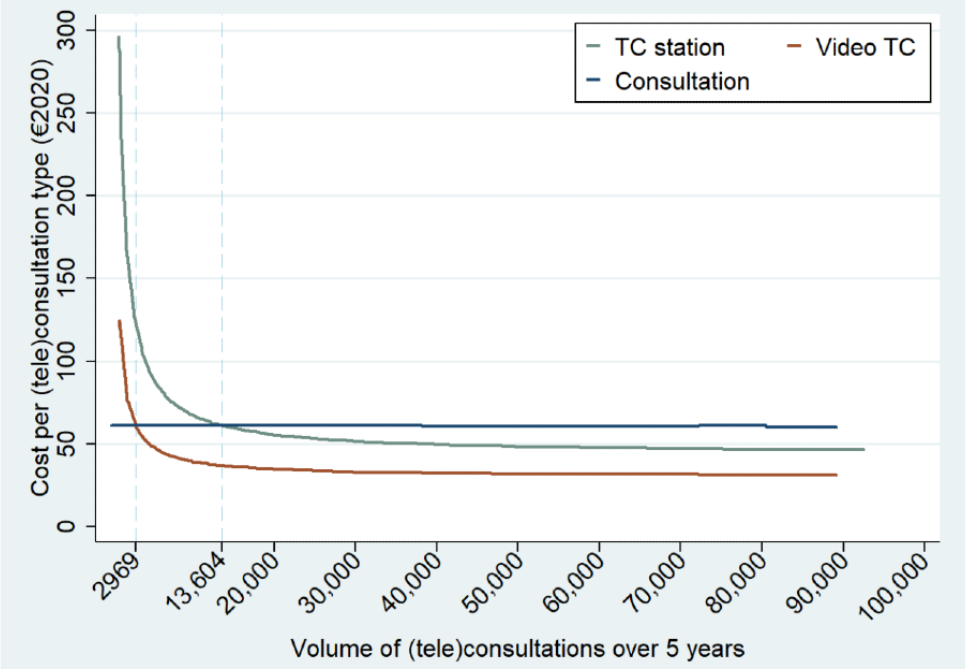


Table 6. Break-even levels for amortization of TC equipment.

Break levels[a]	Visio TC[b]	12 TC stations	Per TC station
Experiment 1 (population level)	2.3/day	10.8/day	0.9/day
Experiment 3 (subpopulation: urban)	3.5/day	28.3/day	2.4/day
Experiment 3 (subpopulation: peri-urban)	2/day	8.8/day	0.7/day
Experiment 3 (subpopulation: remote)	1.2/day	3.3/day	0.3/day
Experiment 2 (subpopulation: nursing homes)	N/A[c]	2.95/day	0.2/day

[a]The break-even point for the amortization of video TC equipment if we restrict the population to urban agents is 3.5 TCs per day over 5 years (253 working days per year). Computation: 4442 video TCs/1265 working days.

[b]TC: teleconsultation.

[c]N/A: not applicable.

In experiments 2 and 3, we replicated experiment 1 within subpopulations to substantiate the potential variation in amortization speed. The subpopulations included elderly people in nursing homes (experiment 2) and people living in urban, peri-urban, and remote areas (experiment 3). Table 6 displays the break-even points for each subpopulation. The videoconferencing solution is amortized if at least 3.5, 2, and 1.2 video TCs per day are completed for an urban, peri-urban, and remote subpopulation, respectively. Thus, the characteristics of the population that composed the market share of the hospital did indeed have a substantial impact on the amortization speed of the videoconferencing solution. The same pattern can be observed for the amortization of telehealth stations, depending on whether the stations are located in urban, peri-urban, and remote areas, with break-even points of 2.4, 0.7, and 0.3 TCs per station per day, respectively. These results substantiate the importance of the location of telehealth stations, which drastically affects the economic profitability of these stations. Finally, our findings indicate that telehealth stations are even more profitable when they are located in nursing homes, with a break-even point of 0.2 TCs per station per day. The economic profitability of telehealth stations was not as straightforward because nursing homes were located relatively close to the hospital (13.06 km on average), which could have increased the break-even point, as is the case for the urban subpopulation. Nevertheless, the fact that nursing home residents are medically transported whenever physical transportation is needed largely offsets the impact of short distances on the total cost.

Finally, the results of the validation experiment are presented in Multimedia Appendix 4, which shows that our simulation strategy worked well in generating a population with characteristics similar to those of the observed population. Similarly, the share of TCs observed in the data is broadly comparable to the simulated share. The volume of (tele)consultations is naturally substantially higher in the simulation than the observed volumes because we considered a 5-year simulation window against a 7-month period of observations for the 2020 data. We could not consider the transfer KPI or cost KPI in the validation experiment because they were not observed in the data.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we conducted an evaluation of the use of TCs for consultations with specialists at a hospital. Using data on (tele)consultations for any eligible specialty, the key feature of our evaluation framework is the combination of an economic evaluation through a cost analysis and a performance evaluation through a DES approach that distinguishes between 2 types of TCs.

Regarding the cost analysis, the use of TCs was found to increase the total cost per patient by €51.19 (US $56) over a 298-day follow-up (Table 5). Naturally, the fixed cost of investment in a videoconferencing solution and in telehealth stations was one of the main driving factors of the incremental cost. Nevertheless, the incremental cost was also found to be driven by a higher number of (tele)consultations per patient (Figure 2). Thus, our results indicate that during the study period, TCs were used as a complement to CSs rather than as a substitute. A probabilistic sensitivity analysis through a bootstrap procedure indicates a strong uncertainty surrounding the point estimation of the incremental cost. Interestingly, the deterministic sensitivity analysis sheds light on the substantial impact of the share of patients with medical transportation on the incremental cost. Indeed, there would no longer be a difference in total cost per patient if 100% of patients were medically transported whenever a CS was needed (Figure 3). The incremental cost was also found to be sensitive to other cost inputs, such as the volume of TCs completed and the number of TC stations.

The sensitivity of the relative cost of TCs compared with CSs suggests a potential heterogeneity in the profitability of this strategy based on the level of deployment (eg, the volume of TCs completed, the number of telehealth stations) and the characteristics of the population (eg, the share of medical transportation). To delve into this heterogeneity, in this paper, we propose a flexible model for evaluating the performance of various scenarios for future TC deployment using a DES approach. Our findings support the notion that the system obtains the worst performance in terms of total cost in scenarios in which there was an investment in telehealth stations or a videoconferencing solution but in which the intensity of use of TCs remained very low (ie, <1% of the eligible population; Figure 4). The poor performance of these scenarios in terms of costs is induced by the substantial initial investment (ie, fixed costs) in a videoconferencing solution or telehealth stations that is spread over a low volume of TCs. By contrast, scenarios relying intensively on video TCs or TCs using telehealth stations were found to be cost saving compared with the scenario without any TCs owing to the presence of substantial scale economies when investing in TC equipment, while CSs are subject to constant returns to scale. Thus, assuming a 5-year life span of TC equipment (ie, 253 working days/year), the videoconferencing solution and a telehealth station would be amortized (ie, break-even point) if at least 2.3 video TCs and 0.9 TCs using a telehealth station per station were completed each day, respectively (Table 6). Our results also substantiate a strong heterogeneity in the economic profitability of each TC type based on the population characteristics. Indeed, the videoconferencing solution would require a break-even point of 3.5 video TCs per day in an urban population, 2 video TCs per day in a peri-urban population, and 1.2 video TCs per day in a remote population. Thus, the amortization speed when investing in a videoconferencing solution is strongly correlated with the target population of the hospital. Similarly, the location of telehealth stations (ie, relative to the hospital location) drastically affects their break-even points, with daily minimum volumes per station of 2.4, 0.7, and 0.3 for TCs using telehealth stations located in urban, peri-urban, and remote areas, respectively. Interestingly, telehealth stations were found to be even more profitable when located in nursing homes, with a break-even point of 0.2 TCs per day per station. Indeed, even if nursing homes were often located relatively close to the hospital (ie, average distance of 13.06 km), the fact that their residents would need medical transportation whenever a CS was needed drastically reinforced the economic profitability of telehealth stations in this context.

Limitations

This study also has several potential limitations. We used a retrospective data set to conduct the economic evaluation, and assignment to the treatment and control groups could not be randomized. To overcome this potential risk of selection bias, we conducted 1:1 matching based on the propensity score to derive a pseudorandomized data set. The configuration of our study (ie, a large control group) was suitable for relying on such matching methods, and the matching showed a good performance in balancing the covariates between our 2 groups of interest (Table 4). We could also not distinguish between video TC and TC using telehealth stations at the individual level in the data because we only observed the share of patients using telehealth stations. We thus had to merge these 2 types of TCs into the same group in the economic evaluation and account for this distinction only in the performance evaluation. Similarly, we lacked other individual data, such as data concerning patient transportation mode and the level of out-of-pocket fees for patients, which we replaced by average data at the national level. Another potential limitation concerns the study period, with data on TCs completed in the early stages of the COVID-19 crisis. The fast adoption of telemedicine during that period, as well as the pressure on the health care system, could have had an impact on the behaviors observed in the data. For example, the COVID-19 crisis might have affected the volume of each TC type observed in the data due to the strategy of delayed medical care adopted by many countries to save medical resources [46]. Nevertheless, in this regard, the strength of this study lies in the fact that it is the first to evaluate the use of TCs and the way in which they were used during the study period through an economic evaluation, as well as the fact that this study explores various scenarios for future deployment through a DES approach. Thus, our goal in the DES approach was specifically to explore various levels of intensity of TC use and to assess their relative performance.

XSL•FO
RenderX

Comparison With Prior Work

In the existing literature, there is no overall consensus regarding the performance of TCs in terms of cost or resource consumption [23,24]. The difficulty of reaching a consensus may partially be explained by the variety of diseases and contexts in which telemedicine has been evaluated. Our results support the notion that the performance of TCs is strongly correlated with the characteristics of the targeted population and the volume of TCs completed. Thus, to reap the full benefit of TC, one needs to identify the most profitable scenarios. In this regard, our results are in line with previous studies and tend to show that the use of TC could be particularly profitable when medical transportation could be avoided [23]. Interestingly, our results also substantiate the claim that telehealth stations could be particularly profitable when located in nursing homes. Indeed, even if nursing homes were located relatively close to the hospital in the case study, the fact that their residents would need medical transportation whenever a CS was needed drastically reinforced the economic profitability of telehealth stations in this context.

Furthermore, while the multitude of diseases investigated independently in a disease-specific setting prevented many studies from considering the amortization of telemedicine equipment, we consider any specialty eligible for TC [6,18-21,23,24]. By so doing, our results substantiate a strong correlation between the break-even levels for TC equipment and the characteristics of the targeted population (ie, their location relative to the hospital). As an illustration, a hospital treating patients living in a radius of 20 km would have to perform 3 times more TCs compared with a hospital treating patients living more than 50 km away to amortize a videoconferencing solution, all else being equal. Similarly, a TC station located close to a hospital (ie, within 20 km) would require 8 times more TCs to be amortized than one located in a remote area (avoiding travel distances >50 km). These findings have important practical implications and substantiate the claim that the strategy of deployment for a TC program should take into account the characteristics of the population and the geographical spread of hospitals and patients within the territory.

Regarding resources other than cost, a recent literature review argued that the use of TCs could provide increased productivity through reduction in consultation time, which might, however, be offset by administrative overhead [23]. In this regard, our results support the notion that the reduction in doctor time resulting from TCs (ie, use of either a videoconferencing solution or a telehealth station) largely offsets the increase in administrative time consumption. Moreover, in the simulation, by distinguishing between video TCs and TCs using telehealth stations, we shed light on another important resource: nurses. When considering nurse time consumption, our findings indicate that the net effect on resource time consumption could result in increased time. However, an increase in resource time consumption is unlikely to affect the total cost from a health care perspective under an activity-based payment scheme. Nevertheless, these variations in resource consumption should be taken into account to calibrate resources properly when introducing a new telehealth station.

Conclusions

To conclude, the use of TCs has the potential to lead to a major organizational change in the health care system in the near future. Nevertheless, the performance of TCs in terms of cost reduction is strongly related to the context and deployment strategy. Decision makers should, for example, pay attention to the volume of TCs they expect to achieve as well as the characteristics of the targeted population when investing in a TC solution because they have a decisive impact on its economic profitability. Furthermore, while the organizational and economic impacts of TCs are 2 major aspects to be taken into account in future TC development, there are several other important aspects not covered in this study that should be taken into account, such as patients' satisfaction and access to care as well as adoption by health care professionals.

Acknowledgments

This work was supported by the Chaire Avenir Santé Numérique, funded by the Fondation de l'Avenir. The authors thank the Clinique Mutualiste de Saint-Etienne for granting access to its database.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Optimal period of follow-up.
[DOCX File , 41 KB - jmir_v24i5e32002_app1.docx]

Multimedia Appendix 2
Description of ICD-10 chapters.
[DOCX File , 14 KB - jmir_v24i5e32002_app2.docx]

Multimedia Appendix 3
Medical patient transportation cost.
[DOCX File , 15 KB - jmir_v24i5e32002_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Validation Experiment.
[DOCX File , 21 KB - jmir_v24i5e32002_app4.docx]

References

1. Weinstein RS, Krupinski EA, Doarn CR. Clinical Examination Component of Telemedicine, Telehealth, mHealth, and Connected Health Medical Practices. Med Clin North Am 2018 May;102(3):533-544. [doi: 10.1016/j.mcna.2018.01.002] [Medline: 29650074]
2. Tedeschi C. Ethical, Legal, and Social Challenges in the Development and Implementation of Disaster Telemedicine. Disaster Med Public Health Prep 2021 Oct 04;15(5):649-656 [FREE Full text] [doi: 10.1017/dmp.2020.118] [Medline: 32364096]
3. Swinton JJ, Robinson WD, Bischoff RJ. Telehealth and rural depression: physician and patient perspectives. Fam Syst Health 2009 Jun;27(2):172-182. [doi: 10.1037/a0016014] [Medline: 19630458]
4. Nesbitt TS, Cole SL, Pellegrino L, Keast P. Rural outreach in home telehealth: assessing challenges and reviewing successes. Telemed J E Health 2006 Apr;12(2):107-113. [doi: 10.1089/tmj.2006.12.107] [Medline: 16620164]
5. French National Health Insurance. Croissance record du recours à la téléconsultations en mars [Communiqué de presse du 31 mars 2020]. de l'Assurance Maladie. 2020 Mar 31. URL: https://assurance-maladie.ameli.fr/sites/default/files/2020-03-31-cp-teleconsultation-croissance-record-mars.pdf [accessed 2022-02-03]
6. French National Authority for Health. Efficience de la télémédecine : état des lieux de la littérature internationale et cadre d'évaluation (Rapport d'évaluation médico-économique). Haute Autorité de Santé. URL: https://www.has-sante.fr/upload/docs/application/pdf/2013-07/efficience_tlm_vf_2013-07-18_14-48-38_743.pdf [accessed 2022-02-03]
7. European Commission. Market study on telemedecine. European Commission. 2018. URL: https://ec.europa.eu/health/system/files/2019-08/2018_provision_marketstudy_telemedicine_en_0.pdf [accessed 2022-02-03]
8. NGAP. Nomenclature Générale des Actes Professionnels. NGAP. 2020. URL: https://www.ameli.fr/sites/default/files/Documents/717399/document/ngap-assurance-maladie-03122020.pdf [accessed 2022-02-03]
9. French National Health Insurance. Avenant N°15 à la convention nationale du 4 avril 2012 organisant les rapports entre les pharmaciens titulaires d'officine et l'assurance maladie. de l'Assurance Maladie. 2018. URL: https://www.ameli.fr/sites/default/files/Documents/633103/document/avenant15_convention_nationale_pharmacien.pdf [accessed 2022-02-03]
10. Buvik A, Bergmo TS, Bugge E, Smaabrekke A, Wilsgaard T, Olsen JA. Cost-Effectiveness of Telemedicine in Remote Orthopedic Consultations: Randomized Controlled Trial. J Med Internet Res 2019 Feb 19;21(2):e11330 [FREE Full text] [doi: 10.2196/11330] [Medline: 30777845]
11. Bergmo TS. Economic evaluation in telemedicine - still room for improvement. J Telemed Telecare 2010 May 25;16(5):229-231. [doi: 10.1258/jtt.2010.009008] [Medline: 20501629]
12. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Quality of care for remote orthopaedic consultations using telemedicine: a randomised controlled trial. BMC Health Serv Res 2016 Sep 08;16:483 [FREE Full text] [doi: 10.1186/s12913-016-1717-7] [Medline: 27608768]
13. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
14. Buvik A, Bugge E, Knutsen G, Småbrekke A, Wilsgaard T. Patient reported outcomes with remote orthopaedic consultations by telemedicine: A randomised controlled trial. J Telemed Telecare 2018 Jul 04;25(8):451-459. [doi: 10.1177/1357633x18783921]
15. Fieux M, Duret S, Bawazeer N, Denoix L, Zaouche S, Tringali S. Téléconsultation en ORL : enquête de satisfaction en période pandémique COVID-19. Annales françaises d'Oto-rhino-laryngologie et de Pathologie Cervico-faciale 2020 Sep;137(4):236-241. [doi: 10.1016/j.aforl.2020.05.012]
16. Ramaswamy A, Yu M, Drangsholt S, Ng E, Culligan PJ, Schlegel PN, et al. Patient Satisfaction With Telemedicine During the COVID-19 Pandemic: Retrospective Cohort Study. J Med Internet Res 2020 Sep 09;22(9):e20786 [FREE Full text] [doi: 10.2196/20786] [Medline: 32810841]
17. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
18. Bergmo TS. Can economic evaluation in telemedicine be trusted? A systematic review of the literature. Cost Eff Resour Alloc 2009 Oct 24;7:18 [FREE Full text] [doi: 10.1186/1478-7547-7-18] [Medline: 19852828]
19. Reardon T. Research findings and strategies for assessing telemedicine costs. Telemed J E Health 2005 Jun;11(3):348-369. [doi: 10.1089/tmj.2005.11.348] [Medline: 16035931]
20. Whitten P, Mair F, Haycox A, May C, Williams T, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [doi: 10.1136/bmj.324.7351.1434] [Medline: 12065269]

XSL•FO
RenderX

21. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
22. French National Authority for Health. Évaluation économique de la télésurveillance pour éclairer la décision publique. Quels sont les choix efficients au regard de l'analyse de la littérature? Haute Autorité de Santé. 2020. URL: https://www.has-sante.fr/upload/docs/application/pdf/2020-12/rapport_evaluation_telesurveillance.pdf [accessed 2022-02-03]
23. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if Telehealth Can Reduce Health System Costs: Scoping Review. J Med Internet Res 2020 Oct 19;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
24. Eze ND, Mateus C, Cravo Oliveira Hashiguchi T. Telemedicine in the OECD: An umbrella review of clinical and cost-effectiveness, patient experience and implementation. PLoS One 2020 Aug 13;15(8):e0237585 [FREE Full text] [doi: 10.1371/journal.pone.0237585] [Medline: 32790752]
25. Fanti MP, Mininel S, Ukovich W, Vatta F. Modelling alarm management workflow in healthcare according to IHE framework by coloured Petri Nets. Engineering Applications of Artificial Intelligence 2012 Jun;25(4):728-733. [doi: 10.1016/j.engappai.2010.11.003]
26. Dotoli M, Fanti MP, Mangini AM, Ukovich W. A continuous Petri net model for the management and design of emergency cardiology departments. Amsterdam, The Netherlands: Elsevier; 2009 Presented at: Third IFAC Conference on Analysis and Design of Hybrid Systems; September 16–18, 2009; Zaragoza, Spain p. 50-55 URL: https://tinyurl.com/pn4698pb [doi: 10.3182/20090916-3-es-3003.00010]
27. Dotoli M, Fanti M, Iacobellis G, Martino L, Moretti A, Ukovich W. Modeling and management of a hospital department via Petri nets. New York, NY: IEEE; 2010 Mar Presented at: 2010 IEEE Workshop on Health Care Management (WHCM); February 18-20, 2010; Venice, Italy p. 1-6 URL: https://www.researchgate.net/publication/224128594_Modeling_and_management_of_a_hospital_department_via_Petri_nets [doi: 10.1109/whcm.2010.5441248]
28. Hamana S, Augusto V, Xie X. A timed Petri net approach for verification of territorial healthcare information systems. New York, NY: IEEE; 2016 Presented at: 2016 IEEE International Conference on Automation Science and Engineering; August 21-24, 2016; Fort Worth, TX. [doi: 10.1109/coase.2016.7743465]
29. Lach J, Vázquez R. Simulation model of the telemedicine program. In: Proceedings of the 2004 Winter Simulation Conference. New York, NY: IEEE; 2004 Presented at: Winter Simulation Conference; December 5-8, 2004; Washington, DC. [doi: 10.1109/wsc.2004.1371563]
30. Qiao Y, Ran L, Li J. Optimization of Teleconsultation Using Discrete-Event Simulation from a Data-Driven Perspective. Telemed J E Health 2020 Jan 01;26(1):114-125. [doi: 10.1089/tmj.2018.0229] [Medline: 30994407]
31. Akuamoah SW, Lu D, Yaro D. Application of Queueing Theory in Dispensation of Resources and Optimization in Teleconsultation. Int. J. Appl. Comput. Math 2020 Jul 14;6(4):1-17. [doi: 10.1007/s40819-020-00851-0]
32. Swan B, Shevlin C, Cho A, Phinney D. Simulation tool to evaluate electronic consultations in rheumatology. New York, NY: IEEE; 2018 Presented at: 2018 Winter Simulation Conference (WSC); December 9-12, 2018; Gothenburg, Sweden p. 2589-2599. [doi: 10.1109/wsc.2018.8632444]
33. Jean C, Jankovic M, Stal-Le Cardinal J, Bocquet J. Predictive modelling of telehealth system deployment. Journal of Simulation 2017 Dec 19;9(2):182-194. [doi: 10.1057/jos.2014.27]
34. French National Authority for Health. Guide Méthodologique: Cartographie Des Impacts Organisationnels Pour l'évaluation Des Technologies de Santé. Haute Autorité de Santé. 2020. URL: https://www.has-sante.fr/upload/docs/application/pdf/2020-12/guide_methodologique_impacts_organisationnels.pdf?msclkid=2e95ee68cf6d11ec90d2b5fd9c9a49cc [accessed 2022-05-08]
35. French National Authority for Health. Methodological Guidance: Choices in Methods for Economic Evaluation. Haute Autorité de Santé. 2020. URL: https://www.has-sante.fr/upload/docs/application/pdf/2020-12/guide_methodologique_impacts_organisationnels.pdf [accessed 2022-02-03]
36. JORF n° 0051 du 29 février 2020. Légifrance. 2020. URL: https://www.legifrance.gouv.fr/jorf/jo/2020/02/29/0051 [accessed 2021-03-03]
37. French National Health Insurance. Données statistiques sur la patientèle des professionnels de santé libéraux. de l'Assurance Maladie. 2017. URL: https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/donnees-statistiques/professionnels-de-sante-liberaux/patientele/patientele-des-medecins.php [accessed 2022-03-29]
38. French National Health Insurance. Tarifs conventionnés des ambulances. de l'Assurance Maladie. 2021. URL: https://www.ameli.fr/transporteur-sanitaire/exercice-professionnel/facturation/tarifs/ambulances-tarifs-conventionnels [accessed 2022-02-03]
39. French National Health Insurance. Tarifs conventionnés des VSL. de l'Assurance Maladie. 2021. URL: https://www.ameli.fr/transporteur-sanitaire/exercice-professionnel/facturation/tarifs/vsl-tarifs-conventionnels [accessed 2022-02-03]
40. French NHI. Facturation : les taxis conventionnés. de l'Assurance Maladie. 2021. URL: https://www.ameli.fr/sites/default/files/annexe4-janvier-2018_cpam-83.pdf [accessed 2022-02-03]
41. Gonzalez L, Héam J, Mikou M, Ferretti C. Les dépenses de santé en 2018 - édition 2019. Résultats des comptes de la santé. Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques. Haute Autorité de Santé. 2019. URL: https://tinyurl.com/4scp24sk [accessed 2022-02-03]

XSL•FO
RenderX

42. Stuart EA. Matching Methods for Causal Inference: A Review and a Look Forward. Statist. Sci 2010 Feb 1;25(1):1-21. [doi: 10.1214/09-sts313]
43. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41-55. [doi: 10.1093/biomet/70.1.41]
44. Efron B. Bootstrap Methods: Another Look at the Jackknife. Ann. Statist 1979 Jan 1;7(1):1-26 [FREE Full text] [doi: 10.1214/aos/1176344552]
45. Gouyon M. Consulter un spécialiste libéral à son cabinet : premiers résultats d'une enquête nationale. Études Et Résultats. URL: https://drees.solidarites-sante.gouv.fr/IMG/pdf/er704.pdf [accessed 2022-02-03]
46. Findling MG, Blendon RJ, Benson JM. Delayed Care with Harmful Health Consequences—Reported Experiences from National Surveys During Coronavirus Disease 2019. JAMA Health Forum 2020 Dec 14;1(12):e201463. [doi: 10.1001/jamahealthforum.2020.1463]

Abbreviations

CS: physical consultation
DES: discrete-event simulation
DREES: Directorate for Research, Studies, Assessment, and Statistics
EQ-5D-3L: EuroQol 5-dimensional system
ICD-10: International Statistical Classification of Disease and Related Health Problems, tenth revision
ICT: information and communications technology
INSEE: National Institute of Statistics and Economic Studies
KPI: key performance indicator
PS: propensity score
TC: teleconsultation
UML: uniform modeling language
VSL: light health vehicle
WMA: World Medical Association

Edited by T Leung; submitted 13.07.21; peer-reviewed by C Jean, H Mehdizadeh; comments to author 02.11.21; revised version received 18.12.21; accepted 07.02.22; published 20.05.22.

Please cite as:
Huguet M, Sarazin M, Perrier L, Augusto V
How We Can Reap the Full Benefit of Teleconsultations: Economic Evaluation Combined With a Performance Evaluation Through a Discrete-Event Simulation
J Med Internet Res 2022;24(5):e32002
URL: https://www.jmir.org/2022/5/e32002
doi:10.2196/32002
PMID:35594065

©Marius Huguet, Marianne Sarazin, Lionel Perrier, Vincent Augusto. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2022. 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.

XSL•FO
RenderX

Original Paper

Health Care Professionals' Experiences of Web-Based Symptom Checkers for Triage: Cross-sectional Survey Study

Sari Kujala[1], PhD; Iiris Hörhammer[2], DSc

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

Corresponding Author:
Sari Kujala, PhD
Department of Computer Science
Aalto University
PO Box 15400
Espoo, FI-00076
Finland
Phone: 358 503862768
Email: sari.kujala@aalto.fi

Abstract

Background: Web-based symptom checkers are promising tools that provide help to patients seeking guidance on health problems. Many health organizations have started using them to enhance triage. Patients use the symptom checker to report their symptoms online and submit the report to the health care center through the system. Health care professionals (registered nurse, practical nurse, general physician, physiotherapist, etc) receive patient inquiries with urgency rating, decide on actions to be taken, and communicate these to the patients. The success of the adoption, however, depends on whether the tools can efficiently support health care professionals' workflow and achieve their support.

Objective: This study explores the factors influencing health care professionals' support for a web-based symptom checker for triage.

Methods: Data were collected through a web-based survey of 639 health care professionals using either of the two most used web-based symptom checkers in the Finnish public primary care. Linear regression models were fitted to study the associations between the study variables and health care professionals' support for the symptom checkers. In addition, the health care professionals' comments collected via survey were qualitatively analyzed to elicit additional insights about the benefits and challenges of the clinical use of symptom checkers.

Results: Results show that the perceived beneficial influence of the symptom checkers on health care professionals' work and the perceived usability of the tools were positively associated with professionals' support. The perceived benefits to patients and organizational support for use were positively associated, and threat to professionals' autonomy was negatively associated with health care professionals' support. These associations were, however, not independent of other factors included in the models. The influences on professionals' work were both positive and negative; the tools streamlined work by providing preliminary information on patients and reduced the number of phone calls, but they also created extra work as the professionals needed to call patients and ask clarifying questions. Managing time between the use of symptom checkers and other tasks was also challenging. Meanwhile, according to health care professionals' experience, the symptom checkers benefited patients as they received help quickly with a lower threshold for care.

Conclusions: The efficient use of symptom checkers for triage requires usable solutions that support health care professionals' work. High-quality information about the patients' conditions and an efficient way of communicating with patients are needed. Using a new eHealth tool also requires that health organizations and teams reorganize their workflows and work distributions to support clinical processes.

(J Med Internet Res 2022;24(5):e33505) doi:10.2196/33505

KEYWORDS

adoption; symptom checker; triage; health care professional; survey; online health; digital health; health organizations; health care

XSL•FO
RenderX

Introduction

Background

Web-based symptom checkers are promising tools that provide help to patients seeking guidance on health problems [1]. Algorithm-assisted symptom checkers ask patients questions about their symptoms and may provide them with potential diagnoses, inform them about the urgency of seeking care, and direct them to appropriate care settings [2].

Many health organizations have started using symptom checkers to guide patients to the most appropriate course of action [2-5]. Notably, evidence of the diagnostic accuracy and impacts of web-based symptom checkers remains scarce [1], but they may supplement resource-intensive telephone triage lines common in primary care [2]. Recently, the COVID-19 pandemic has created pressure to use web-based symptom checkers to avoid face-to-face contact and preserve urgent care capacity [6,7].

Patients are generally highly satisfied with symptom checkers [1]. In a survey study involving the Isabel Symptom Checker, most patients perceived it to be useful for diagnosis [8]. They often used the symptom checker to better understand the reason for their symptoms and to decide whether to seek in-person health care. Users of the Erdusyk Symptom Checker also acknowledged its usefulness to avoid unnecessary visits to general practitioners [3].

However, the impact of symptom checkers on the urgent care system and the wider scope of health care is not known [1]. For example, triage advice is generally risk averse, encouraging patients to seek care for conditions in which self-care would be more appropriate [2]. In a recent study, only 5 out of 15 symptom checkers were found to be superior to the accuracy of laypersons, and the services were suspected to increase resource use in health care [9]. Moreover, symptom checkers are most often used by younger, more educated, and female patients [1,6].

As technology develops, the performance of symptom checkers can be expected to improve [2]. However, their success depends on health care professionals' acceptance and the influence of symptom checkers on their workflow [10]. Health care professionals also influence patients' adoption of eHealth services with their endorsement [11,12].

The goal of this survey study is to identify health care professionals' experiences of symptom checker use in triage and analyze factors influencing their support for these tools. The two examined symptom checkers are used to help patients in obtaining appropriate care, but only one of them provides the patient with immediate information on conditions that correspond to their symptoms. Patients report their symptoms online and submit the report to the health care center through the symptom checker. Health care professionals (registered nurse, practical nurse, general physician, physiotherapist, etc) receive patient inquiries with urgency rating, decide on actions to be taken, and communicate these to the patients. The findings provide a better understanding of how symptom checkers can support efficient clinical work. Sustainable use of eHealth tools requires engagement of both professionals and patients [13]. Therefore, we also study whether professionals' support is manifested in their interactions with patients.

Research Model

The research model is based on the previous version that was tested in the preimplementation phase of a patient portal [14]. Similar to what was found in the preimplementation phase, we hypothesize the following notions after implementation, when the symptom checkers are used in health care organizations:

Hypothesis 1: the perceived usability of the symptom checker is positively associated with health care professionals' support for the symptom checker.

Hypothesis 2: the positive influences on the work are positively associated with health care professionals' support for a symptom checker.

Hypothesis 3: the positive influences on patients are positively associated with health care professionals' support for a symptom checker.

Hypothesis 4: the perceived threat to professional autonomy is negatively associated with health care professionals' support for a symptom checker.

In the preimplementation phase, expectations of good implementation practices in the work unit were positively associated with professionals' support [14]. Accordingly, we hypothesize the following:

Hypothesis 5: organizational support for the use of the symptom checker is positively associated with professionals' support for a symptom checker.

Professionals' endorsement of eHealth tools is crucial in a patient's decision to adopt health technology [15]. To test whether professionals' support is manifested in their interactions with the patients, we hypothesize the following:

Hypothesis 6: health care professionals' support for a symptom checker is positively associated with professionals' promotion of the tool to the patients.

Hypothesis 7: health care professionals' support for a symptom checker is positively associated with their provision of instruction to patients on the use of the tool.

Methods

This study was designed as a cross-sectional survey of the health care professionals using symptom checkers for triage in Finland. A web-based questionnaire was developed to identify the experiences of health care professionals, including nurses, physiotherapists, and physicians.

Study Setting

Both web-based symptom checkers studied have been adopted in Finnish public primary care. They are used to help patients in obtaining appropriate care. Inquiries coming into primary care can be digitally managed through the services instead of traditional phone-based triage. Patients report their symptoms online, and if they wish, they submit a report to health centers. The symptom checkers assess the urgency of care, and, if

XSL•FO
RenderX

needed, guide patients to contact emergency care. Health care professionals (usually nurses) receive a list of inquiries with urgency rating to decide on actions to be taken and communicate these to the patients. Professionals may also inquire further information from patients or forward the task to a physician or another professional if needed. Detailed information on the symptom checkers following the guideline extension for evaluation of interventions with an artificial intelligence component [16] is in Multimedia Appendix 1.

In the first symptom checker, Omaolo oirearvio, the patient first chooses a symptom checker that best describes their condition, such as low-back pain, urinary tract infection, or upper respiratory tract infection. In March 2020, the number of the symptom checkers had increased to 16, including a generic symptom checker that is not specific to any certain symptoms. While diagnosis can only be made by a medical doctor in Finland, Omaolo oirearvio provides information on health problems that correspond to the patient-reported symptoms and recommendations for self-care. Professionals evaluate the patient's inquiry by reading a provided summary of the symptoms, relate them to the patient's medical records, and contact the patient using the messaging functionality. Unlike the other symptom checkers, the triage process of the urinary tract infection and COVID-19 symptom checkers can be entirely digitalized. A patient can reserve a COVID-19 test or reserve a prescription for urinary tract infection using the symptom checkers.

Omaolo oirearvio has been evaluated to be easy to use and understandable to most patients [17]. At the beginning of March 2020, Omaolo oirearvio was adopted in 79 municipalities, including the largest cities in Finland. In 2020, the use by health care provider organizations was increased to cover 69% of the Finnish population. The average number of patient users was 120,000 per week from March to September 2020. A total of 1,937,469 responses were recorded into the most frequently used Omaolo COVID-19 symptom checker in 2020 [18].

The second symptom checker, called Klinik Access, is a generic tool with a visualization of human body on the starting page. The patient first chooses the locus of the symptom and then proceeds with the reductionary dynamic form, which adapts the selections on each page for spawning the next set of possible responses onto the next page. Health care professionals receive all information gathered from the patient and an inquiry summary including preliminary diagnoses and urgency estimates. For the health care provider organizations, the tool allows symptom checking and urgency assessment to prioritize patient care [5]. The tool was adopted in 26 municipalities and private health care provider organizations in 2016-2019. The average number of patient users was 33,000 per week in 2020.

Questionnaire

Two earlier surveys were used as models for the current study. The first identified health care professionals' expectations of the patient portal Omaolo, which also includes oirearvio symptom checker [19]. The second survey focused on the health care professionals' first experiences with Omaolo oirearvio [20]. In this new version of the questionnaire, health care professionals were asked to evaluate their experiences with Omaolo oirearvio or Klinik Access symptom checkers (Multimedia Appendix 2).

The questionnaire included existing validated survey items for measuring health care professionals' support for the symptom checkers [21-23], their usability [24], and their influence on professional autonomy [25]. In addition, participants were asked to rate whether the symptom checkers had brought the planned benefits to their work and to patients. The use of these measures was piloted in the previous survey studies [14,20]. Two statements were made regarding professionals' endorsement of the tool to patients: how often they had (1) recommended the use of and (2) instructed patients in the use of the symptom checker. For analysis, these variables were recoded into binary variables (never or at least once). Moreover, 2 open-ended questions were posed in relation to the benefits and challenges brought about by the symptom checkers to respondent's unit or own work, which are as follows. (1) In your opinion, what challenges does the symptom checker bring to your unit or to your work? (2) In your opinion, what benefits does the symptom checker provide to your unit or your own work? To elicit background information, participants were asked about their age, gender, profession, how often they had used the symptom checkers during the previous month, and whether they had participated in the planning of the symptom checkers.

Data Gathering

Data were gathered from February to September 2020 after the symptom checkers had been used in the 36 health organizations from 10 to 32 months. The project manager of each of the organizations sent the questionnaire link via email to the health care professionals who used Omaolo oirearvio. A manager of Klinik Healthcare Solutions, which developed Klinik Access, sent the survey invitation to the health care organizations using it. To encourage participation, 50 movie tickets for Oirearvio users and 3 tablet computers for Klinik Access users were raffled off.

Ethics Approval

The study protocol was reviewed and approved by Aalto University Research Ethics Committee (reference 95_03.04_2019_DigiIN).

Data Analysis

Statistical analysis was performed for the quantitative data. Descriptive statistics and reliability analyses were performed for all study variables (Multimedia Appendix 3). The Cronbach alpha scores for the scales were all above .75 indicating good internal consistency [26]. We fitted an ordinary least squares regression with robust standard errors to study the association of key independent variables, namely, benefits to professionals' work, threat to autonomy, benefits to patients, usability, and organizational support for use with the dependent variable, namely, professionals' support for the symptom checker. First, we fitted the univariate analyses for each independent variable to check its association with the dependent variable. Second, a model was formulated with all key variables as independent variables. Third, we added adjustments for age, gender, the symptom checker solution used, profession, participation in planning, and frequency of use. In addition, logistic regression

XSL•FO
RenderX

models were fitted to study the association of the professionals' support for the symptom checker with their promotion of it to patients and their instruction of patients in its use. These models were also adjusted for age, gender, the symptom checker solution used, profession, participation in planning, and frequency of use. In all analyses, *P* values below the .05 threshold were considered significant. When fitting the multivariate models, independent variables were added simultaneously. To test for multicollinearity, we calculated the variance inflation factors for independent variables. They were all below 2.5, indicating that multicollinearity was not a concern in this study [27].

Qualitative data were content analyzed using Atlas.ti analysis tool. Open coding was used to identify themes in the data without predefined categories. Using in vivo coding, the respondents' words were used to define the themes to ensure that they represent the original intended meaning of the respondents.

Results

Respondents

The characteristics of the respondents are presented in Table 1. The respondents well represented Finnish health care professionals in terms of age, gender, and profession [28]. For example, the mean age in our sample was 42.7 years (43.0 in eligible population in Finland), and the proportion of female participants was 90.7% (88.0% in Finland). Doctors were underrepresented in comparison to nurses; in our sample, there was a little more than 1 doctor per 10 nurses, while in 2014, this number was 2.5 doctors per 10 nurses in Finland. This may be due to the large number of nurses as direct users of the symptom checkers.

Table 1. Respondent characteristics (n=639).

Characteristics	Values
Age (years), mean (SD)	42.7 (11.5)
Gender, n (%)	
Woman	577 (90.3)
Man	41 (6.4)
Other or not reported	21 (3.3)
Profession, n (%)	
Nurse, public nurse, or practical nurse	477 (74.7)
Doctor	54 (8.5)
Physiotherapist	41 (6.4)
Other	67 (10.5)
Participated in planning, n (%)	
No	568 (88.9)
Yes	58 (9.1)
Does not know or not reported	13 (2)
Has provided feedback, n (%)	
No	400 (62.6)
Yes	225 (35.2)
Does not know or not reported	14 (2.2)
Frequency of use, n (%)	
Every day during the last month	209 (32.7)
Every week during the last month	244 (38.2)
1-2 times during the last month	97 (15.2)
Less than monthly but have tried	75 (11.7)
Never used	14 (2.2)

Factors Influencing Professionals' Support for a Symptom Checker

Table 2 presents the results of the linear regression analyses testing the association of independent variables with professionals' support for the symptom checkers. In the univariate analysis (not shown in the table), all key variables were associated with professionals' support. All associations were positive, except for the association of threat to autonomy with support, which was negative. When added to a multivariate

XSL•FO
RenderX

model (Model A), the associations of benefits to professionals' work, usability, and benefits to patients with professionals' support for the symptom checker remain, while the associations of organizational support for use and threat to autonomy disappear.

The independent variables account for 52% of the variation in support. The statistically significant association of benefits to patients with support disappears when adjustments for age, gender, profession, participation in planning, and frequency of use (Model B) are made. The associations of benefits to professionals' work and usability with support remain after adjustments. Thus, the results support hypotheses 1 and 2, but only partially support hypothesis 3. No support was found for hypotheses 4 and 5. Age, gender, and participation in planning were not associated with professionals' support.

Table 2. Regression model results—association of independent variables with professionals' support. Robust standard errors were used. Continuous variables were used as continuous standardized variables.

Variable	Model A		Model B	
	ß (SE)	*P* value	*ß* (SE)	*P* value
Benefits to professionals' work	.37 (.06)	<.001	.39 (.07)	<.001
Threat to autonomy	–.02 (.03)	.46	–.03 (.03)	.34
Benefits to patients	.08 (.04)	.04	.06 (.04)	.16
Usability	.27 (.05)	<.001	.27 (.05)	<.001
Organizational support for use	.02 (.03)	.55	.02 (.03)	.52
Age	N/A[a]	N/A	.00 (.00)	.18
Gender (category reference: woman)	N/A	N/A	–.06 (.10)	.55
Solution (category reference: Klinik)	N/A	N/A	–.02 (.08)	.81
Profession (category reference: nurse, midwife, or public health nurse)	N/A	N/A	N/A	N/A
Doctor	N/A	N/A	.27 (.09)	.004
Physiotherapist	N/A	N/A	.09 (.12)	.44
Other	N/A	N/A	–.09 (.11)	.40
Participated in planning (category reference: yes)	N/A	N/A	–.06 (.11)	.56
Frequency of use (category reference: every day during the last month)	N/A	N/A	N/A	N/A
Every week during the last month		N/A	–13 (.07)	.047
1-2 times during the last month	N/A	N/A	–.10 (.09)	.27
Less than monthly but have tried	N/A	N/A	–.06 (.10)	.54
Never used	N/A	N/A	–.20 (.17)	.26
R-squared	.52	N/A	.52	N/A

[a]N/A: not applicable.

Association of Professionals' Support With Their Recommendation of the Tool to Patients and Provision of Instruction Regarding Its Use

The results of the logistic regression models (Multimedia Appendices 4 and 5) show that support for the symptom checkers was associated with both professionals' recommendation to the patients and their provision of instruction to patients regarding use of the symptom checkers. The associations were maintained even when adjustments for gender, age, profession, symptom checker solution, and participation in the planning of the solution were made. Hypotheses 6 and 7 were thus supported.

Perceived Benefits and Disadvantages of the Symptom Checkers

Of the total 639 health care professionals, 216 (33.8%) responded to the open-ended questions. Most of the health care professionals (164/216, 75.9%) reported both benefits and disadvantages of the symptom checkers, 30/216 (13.9%) perceived only disadvantages, and 22/216 (10.2%) perceived only benefits.

Table 3 summarizes the benefits of the symptom checkers that health care professionals described in their responses to the open-ended questions. The symptom checkers were perceived to be beneficial for work, as they streamline the work in various ways, such as by providing preliminary information on patients and by decreasing telephone work. Patients were able to receive help expeditiously by using symptom checkers. The symptom

XSL•FO
RenderX

checkers were considered to lower the threshold for care, and the self-care instructions were perceived useful for patients.

The most frequently mentioned challenges were related to the characteristics and use of the symptom checkers (Table 4). The results of the assessment provided by the symptom checkers were perceived to be inaccurate and unfocused, so the professionals still needed to call patients and ask clarifying questions.

Users of the symptom checkers found it challenging to use the services alongside other work tasks, as switching between tasks is troublesome. Patients contacted them using several channels, which were a symptom checker, calling, and visiting. Evidently, the symptom checkers did not clearly inform patients how a health care professional will contact them.

Many also mentioned that patients cannot use the symptom checker, or that they do not understand the questions or wordings. Health care professionals were also concerned that older patients are unable to use the service. They suggested that the services should be advertised more, as patient usage is low.

Table 3. Perceived benefits of the symptom checkers evaluated.

Themes	Mentions, n (%)
Facilitates health care professionals' work	
Streamlines work by providing preliminary information on patients	75 (11.7)
Reduces the number of phone calls	52 (8.1)
Makes work flexible, a patient case can be handled at a suitable time slot	27 (4.2)
Comprehensive preliminary information on patients	12 (1.9)
Patients' own descriptions of the symptoms can be used in medical reports	5 (0.8)
Fluent communication with patients	5 (0.8)
Gives variety to the work	5 (0.8)
Reduce the number of visits	5 (0.8)
The symptom checker is beneficial for patients	
Patients receive help easily and quickly	49 (7.7)
Supports self-management	22 (3.4)
Lowers threshold for care	20 (3.1)
Urinary tract infection and sexually transmitted disease symptom checkers are especially useful	19 (3.0)
Uniforms quality of triage	18 (2.8)
Useful during the COVID-19 pandemic	8 (1.3)

XSL•FO
RenderX

Table 4. Perceived disadvantages of the symptom checkers evaluated.

Themes	Mentions, n (%)
The symptom checker is not working in an optimal way	
Communicating with patients is time-consuming or cumbersome	74 (11.6)
There is a need to call patients and ask clarifying questions	56 (8.8)
Provides inaccurate results	54 (8.5)
Cumbersome to use or it should be more automatic	49 (7.7)
Not interoperable with patient health records	36 (5.6)
Too sensitive	19 (3.0)
Provides a poor summary; it is difficult to identify essentials	16 (2.5)
It is easier and faster to evaluate a patient's condition (eg, breathing over the phone)	15 (2.3)
Appointments are challenging to make via the symptom checker	14 (2.2)
Signing in repeatedly is slow	13 (2.0)
The workflow with the symptom checker is not optimal	
Is included among many other tasks, and managing time between different tasks is challenging	66 (10.3)
Creates extra work or slows down working	53 (8.3)
The method for organizing work is unclear; a commonly agreed course of action for responding to patients is missing	17 (2.7)
Patients contact using several channels	31 (4.9)
Patients have difficulties with the symptom checker	
Not all patients are able to use symptom checkers (eg, older people)	33 (5.2)
Few patients use symptom checkers; there should be more advertising	30 (4.7)
Patients do not know how to use the symptom checker	14 (2.2)
Patients do not understand all the questions or wordings of the service	10 (1.6)
It is not clear for patients how professionals contact them	2 (0.3)
Health care professionals need more experience or do not know how to use the symptom checker	8 (1.3)
Resistance to change	7 (1.1)

Discussion

Principal Findings

The results indicate that in the sustained use of the symptom checkers for triage, benefits to health care professionals' work, and usability were the most important factors influencing professionals' support for the tool. Benefits to patients were also positively associated with health care professionals' support for symptom checkers as has previously been found in the implementation phase [14]. However, the association weakened when the control variables were added to the model.

Organizational support for use was positively associated, and threat to professionals' autonomy was negatively associated with health care professionals' support. These associations were, however, not independent of professionals' perceptions of usability and benefits to work. Organizational support may have already been incorporated to the perception of the technology's usability and benefits in work [29].

The open-ended question responses of the professionals deepen understanding of the underlying reasons for the associations found in the quantitative analysis. The influences on professionals' work were seen both positive and negative; the tools streamlined work by providing preliminary information on patients and reduced the number of phone calls, but they also created extra work as the professionals needed to call patients and ask clarifying questions. The need to ask more questions arose from the experienced inaccuracy of the symptom checkers and unclear assessment reports. Health care professionals also reported that patients do not always know how to use symptom checkers, or they may not understand all the wordings of the service.

Based on the health care professionals' experiences, symptom checkers are beneficial for patients as they receive help quickly with lower threshold. Moreover, the services support self-management. The experienced benefits of symptom checkers of sensitive diseases, such as sexually transmitted diseases, suggest that one of the potentials of the symptom checkers may be to lower the threshold for care in such cases, as suggested by Johansen et al [30]. The perceived uniform quality of triage is also beneficial to patients as all essential questions are asked regardless of the skill level of the health care professional. Symptom checkers were perceived to be particularly beneficial during the COVID-19 epidemic. In fact, during the second wave, 1,550,000 people used the Omaolo COVID-19 symptom checker [18].

XSL•FO
RenderX

Comparison With Prior Work

Consistent with earlier studies that focused on professionals' attitudes, usability and utility were the most common factors in promoting adoption by health care professionals [14,31,32]. For symptom checkers, their effective utility in professionals' work was critical. This is in line with earlier findings that the fit between clinical work tasks and the technology design has a significant influence on the adoption of health innovations [10,33]. In agreement with our results, Entezarjou et al [34] found that automated patient interview can streamline clinical work. However, our results support an earlier interpretation that intelligent triage tools may also increase professionals' workload, as the information provided by patients via these tools entails gaps and uncertainty in data [35].

In line with this study, Cajander et al [36] found that digital communication with patients may in some cases slow down the assessment of care need. However, in their study, nurses found digital communication to be emotionally less stressful than phone calls. Better care for patients has been found to be an important benefit of eHealth services from professionals' [34,37] and leaders' point of view [38]; nonetheless, in the acceptance models, benefits to patients have been overlooked [39].

Professionals' experience of the inaccuracy of the assessment reports may have emerged from many sources, such as how they perceived patients' use of symptom checkers. As Marco-Ruiz et al [3] mentioned, the accuracy of the symptom checkers depends on how well patients are able to communicate their symptoms with the tools. Based on their interview study, Tsai et al [40] found that patients sought explanations for the results obtained from web-based symptom checkers. They showed that better explanations and more transparent results improved patient trust on the diagnostic quality of the results and helped them come up with better decisions.

Limitations

The limitations of this study are related to the cross-sectional, single-informant design, as well as the omission of likely relevant contextual factors in the models. First, this study relied on professionals' self-reported perceptions of symptom checkers. The single-informant design may lead to inflation of the strength of relationships. To mitigate this problem, the measures validated in previous studies were applied. Second, in this study, we were not able to perform a longitudinal analysis of the implementation of the symptom checkers in the organizations. To address the dynamic nature of information technology implementation, we compared our findings with professionals' perceptions of another eHealth tool studied in the preimplementation phase in Finland. The findings allowed us to come up with preliminary suggestions on which factors explain professionals' support for an eHealth tool in the preimplementation phase as opposed to the use phase. Third, to limit the length of the questionnaire, we were not able to include all relevant contextual factors in our analysis. The factors that may offer support for symptom checkers are, for instance, the competence of the professionals in using these tools [13,41] and vision for the future development of the tools [19], among others. However, as various symptom checkers are used across many health care contexts and numerous purposes, the research results may not be generalizable to other settings.

Future Directions for Research and Practice

The findings suggest that more research is needed on the successful blended use of digital and traditional communication channels by professionals in triage. In addition, many health care professionals were concerned that not all patients, particularly the older people, are able to use symptom checkers; usage statistics confirm that 20- to 29-year-olds were the largest user group [18]. Thus, future studies are needed to support the wider adoption of symptom checkers and health equity. Furthermore, the transparency and explainability of symptom checkers are worth studying in the future from health care professionals' point of view.

The results imply that it is imperative to develop symptom checkers that are usable and support health care professionals' work. Furthermore, health organizations and teams need to carefully reorganize the work processes and distribution of work so that the use of symptom checkers is smoothly integrated among other tasks. For example, the phone and digital work can be allocated to health care professionals every other week. In addition, health organizations need to ensure that patients are well instructed, are aware that the professionals will contact them, and know that they do not need to initiate contact using several channels simultaneously.

While this research did not investigate the specific tasks different health care professionals perform using the symptom checkers, it is likely that the tasks in the initial and more specialized screening of patient inquiries differ between professional groups. To better understand the potential of symptom checkers to supplement resource-intensive telephone triage lines common in primary care, more research on division of tasks between different professional groups in symptom checker–supported triage is needed.

Acknowledgments

This work was supported by the Strategic Research Council at the Academy of Finland (grant numbers 327145 and 327147) and NordForsk through the funding to Nordic eHealth for Patients: Benchmarking and Developing for the Future (NORDeHEALTH), project number 100477.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Description of symptom checkers.
[DOCX File , 1472 KB - jmir_v24i5e33505_app1.docx]

Multimedia Appendix 2
Questionnaire.
[DOCX File , 16 KB - jmir_v24i5e33505_app2.docx]

Multimedia Appendix 3
The means, standard deviations, and Cronbach alphas of key study variables.
[DOCX File , 14 KB - jmir_v24i5e33505_app3.docx]

Multimedia Appendix 4
Logistic regression results; predictors of promotion of the symptom checker to the patients.
[DOCX File , 17 KB - jmir_v24i5e33505_app4.docx]

Multimedia Appendix 5
Logistic regression results; predictors of instructing patients in the use of the symptom checker.
[DOCX File , 17 KB - jmir_v24i5e33505_app5.docx]

References

1. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
2. Semigran H, Linder J, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
3. Marco-Ruiz L, Bønes E, de la Asunción E, Gabarron E, Aviles-Solis JC, Lee E, et al. Combining multivariate statistics and the think-aloud protocol to assess Human-Computer Interaction barriers in symptom checkers. J Biomed Inform 2017 Oct;74:104-122 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.002] [Medline: 28893671]
4. Miller S, Gilbert S, Virani V, Wicks P. Patients' Utilization and Perception of an Artificial Intelligence-Based Symptom Assessment and Advice Technology in a British Primary Care Waiting Room: Exploratory Pilot Study. JMIR Hum Factors 2020 Jul 10;7(3):e19713 [FREE Full text] [doi: 10.2196/19713] [Medline: 32540836]
5. Tenhunen H, Hirvonen P, Linna M, Halminen O, Hörhammer I. Intelligent Patient Flow Management System at a Primary Healthcare Center - The Effect on Service Use and Costs. Stud Health Technol Inform 2018;255:142-146. [Medline: 30306924]
6. Morse KE, Ostberg NP, Jones VG, Chan AS. Use Characteristics and Triage Acuity of a Digital Symptom Checker in a Large Integrated Health System: Population-Based Descriptive Study. J Med Internet Res 2020 Nov 30;22(11):e20549 [FREE Full text] [doi: 10.2196/20549] [Medline: 33170799]
7. Perlman A, Vodonos Zilberg A, Bak P, Dreyfuss M, Leventer-Roberts M, Vurembrand Y, et al. Characteristics and Symptoms of App Users Seeking COVID-19-Related Digital Health Information and Remote Services: Retrospective Cohort Study. J Med Internet Res 2020 Oct 20;22(10):e23197 [FREE Full text] [doi: 10.2196/23197] [Medline: 32961527]
8. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient Perspectives on the Usefulness of an Artificial Intelligence-Assisted Symptom Checker: Cross-Sectional Survey Study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
9. Schmieding ML, Mörgeli R, Schmieding MAL, Feufel MA, Balzer F. Benchmarking Triage Capability of Symptom Checkers Against That of Medical Laypersons: Survey Study. J Med Internet Res 2021 Mar 10;23(3):e24475. [doi: 10.2196/24475]
10. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
11. Wald J. Variations in patient portal adoption in four primary care practices. AMIA Annu Symp Proc 2010 Nov 13;2010:837-841 [FREE Full text] [Medline: 21347096]
12. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
13. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]

XSL•FO
RenderX

14. Kujala S, Hörhammer I, Kaipio J, Heponiemi T. Health professionals' expectations of a national patient portal for self-management. Int J Med Inform 2018 Sep;117:82-87 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.06.005] [Medline: 30032968]
15. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013 Apr 04;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
16. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK, SPIRIT-AICONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. BMJ 2020 Sep 09;370:m3164 [FREE Full text] [doi: 10.1136/bmj.m3164] [Medline: 32909959]
17. Liu VD, Sellgren L, Kaila M, Koskela T. Sähköisten oirearvioiden käytettävyys. FinJeHeW 2021 Jun 18;13(2):100-112. [doi: 10.23996/fjhw.97020]
18. Jormanainen V, Soininen L. Use and Users of the Web-Based Omaolo Covid-19 Symptom Self-Assesment Tool in Finland Since March 16, 2020. In: Studies in Health Technology and Informatics. Amsterdam, The Netherlands: IOS Press; 2021:739-743.
19. Kujala S, Hörhammer I, Heponiemi T, Josefsson K. The Role of Frontline Leaders in Building Health Professional Support for a New Patient Portal: Survey Study. J Med Internet Res 2019 Mar 22;21(3):e11413 [FREE Full text] [doi: 10.2196/11413] [Medline: 30901003]
20. Kujala S, Hörhammer I, Hänninen-Ervasti R, Heponiemi T. Health Professionals' Experiences of the Benefits and Challenges of Online Symptom Checkers. Stud Health Technol Inform 2020 Jun 16;270:966-970. [doi: 10.3233/SHTI200305] [Medline: 32570525]
21. Ward MM, Vartak S, Schwichtenberg T, Wakefield DS. Nurses' perceptions of how clinical information system implementation affects workflow and patient care. Comput Inform Nurs 2011 Sep;29(9):502-511. [doi: 10.1097/NCN.0b013e31822b8798] [Medline: 21825972]
22. Wakefield DS, Halbesleben JRB, Ward MM, Qiu Q, Brokel J, Crandall D. Development of a measure of clinical information systems expectations and experiences. Med Care 2007 Sep;45(9):884-890. [doi: 10.1097/MLR.0b013e3180653625] [Medline: 17712259]
23. Kirkendall ES, Goldenhar LM, Simon JL, Wheeler DS, Andrew Spooner S. Transitioning from a computerized provider order entry and paper documentation system to an electronic health record: expectations and experiences of hospital staff. Int J Med Inform 2013 Nov;82(11):1037-1045. [doi: 10.1016/j.ijmedinf.2013.08.005] [Medline: 24041453]
24. Finstad K. The Usability Metric for User Experience. Interacting with Computers 2010 Sep;22(5):323-327. [doi: 10.1016/j.intcom.2010.04.004]
25. Walter Z, Lopez MS. Physician acceptance of information technologies: Role of perceived threat to professional autonomy. Decision Support Systems 2008 Dec;46(1):206-215. [doi: 10.1016/j.dss.2008.06.004]
26. Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 1981 Feb;18(1):39. [doi: 10.2307/3151312]
27. O'brien RM. A Caution Regarding Rules of Thumb for Variance Inflation Factors. Qual Quant 2007 Mar 13;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
28. Virtanen A. Terveys- ja sosiaalipalveluiden henkilöstö 2014 [Health care and social welfare personnel 2014]. National Institute for Health and Welfare. 2018. URL: http://www.julkari.fi/handle/10024/135915 [accessed 2022-04-25]
29. Leonardi PM. When Flexible Routines Meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies. MIS Quarterly 2011;35(1):147. [doi: 10.2307/23043493]
30. Johansen MA, Berntsen GKR, Schuster T, Henriksen E, Horsch A. Electronic symptom reporting between patient and provider for improved health care service quality: a systematic review of randomized controlled trials. part 2: methodological quality and effects. J Med Internet Res 2012 Oct 03;14(5):e126 [FREE Full text] [doi: 10.2196/jmir.2216] [Medline: 23032363]
31. Gagnon M, 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 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
32. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
33. Ammenwerth E. Technology Acceptance Models in Health Informatics: TAM and UTAUT. Stud Health Technol Inform 2019 Jul 30;263:64-71. [doi: 10.3233/SHTI190111] [Medline: 31411153]
34. Entezarjou A, Bolmsjö BB, Calling S, Midlöv P, Milos Nymberg V. Experiences of digital communication with automated patient interviews and asynchronous chat in Swedish primary care: a qualitative study. BMJ Open 2020 Jul 23;10(7):e036585 [FREE Full text] [doi: 10.1136/bmjopen-2019-036585] [Medline: 32709650]
35. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May 07;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]

XSL•FO
RenderX

36. Cajander Å, Larusdottir M, Hedström G. The effects of automation of a patient-centric service in primary care on the work engagement and exhaustion of nurses. Qual User Exp 2020 Sep 19;5(1). [doi: 10.1007/s41233-020-00038-x]
37. Miller DP, Latulipe C, Melius KA, Quandt SA, Arcury TA. Primary Care Providers' Views of Patient Portals: Interview Study of Perceived Benefits and Consequences. J Med Internet Res 2016 Jan 15;18(1):e8 [FREE Full text] [doi: 10.2196/jmir.4953] [Medline: 26772771]
38. Laukka E, Kujala S, Gluschkoff K, Kanste O, Hörhammer I, Heponiemi T. Leaders' support for using online symptom checkers in Finnish primary care: Survey study. Health Informatics J 2021 Nov 25;27(4):14604582211052259 [FREE Full text] [doi: 10.1177/14604582211052259] [Medline: 34821152]
39. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health Care Provider Adoption of eHealth: Systematic Literature Review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
40. Tsai C, You Y, Gui X, Kou Y, Carroll J. Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers. 2021 Presented at: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems; May 2021; Yokohama, Japan p. 1-17. [doi: 10.1145/3411764.3445101]
41. Kujala S, Rajalahti E, Heponiemi T, Hilama P. Health Professionals' Expanding eHealth Competences for Supporting Patients' Self-Management. Stud Health Technol Inform 2018;247:181-185. [Medline: 29677947]

Edited by T Leung; submitted 10.09.21; peer-reviewed by A Meyer, S Gilbert; comments to author 14.10.21; revised version received 27.10.21; accepted 31.03.22; published 05.05.22.

Please cite as:
Kujala S, Hörhammer I
Health Care Professionals' Experiences of Web-Based Symptom Checkers for Triage: Cross-sectional Survey Study
J Med Internet Res 2022;24(5):e33505
URL: https://www.jmir.org/2022/5/e33505
doi:10.2196/33505
PMID:35511254

©Sari Kujala, Iiris Hörhammer. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2022. 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.

XSL•FO
RenderX

Original Paper

Promised and Lottery Airtime Incentives to Improve Interactive Voice Response Survey Participation Among Adults in Bangladesh and Uganda: Randomized Controlled Trial

Dustin Garrett Gibson[1], MS, PhD; Gulam Muhammed Al Kibria[1], MBBS, MSPH; George William Pariyo[1], MBChB, PhD; Saifuddin Ahmed[1], MBBS, PhD; Joseph Ali[1], JD; Alain Bernard Labrique[1], MS, MHS, PhD; Iqbal Ansary Khan[2], MBBS, MSc, MPH; Elizeus Rutebemberwa[3], MBChB, MPH, PhD; Meerjady Sabrina Flora[2], MBBS, MPH, PhD; Adnan Ali Hyder[4], MD, MPH, PhD

[1]Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

[2]Institute of Epidemiology, Disease Control and Research, Dhaka, Bangladesh

[3]Makerere University School of Public Health, Makerere University College of Health Sciences, Kampala, Uganda

[4]Milken Institute School of Public Health, George Washington University, Washington DC, MD, United States

Corresponding Author:
Dustin Garrett Gibson, MS, PhD
Johns Hopkins Bloomberg School of Public Health
615 N Wolfe Street
E8650
Baltimore, MD, 21205
United States
Phone: 1 443 287 8763
Email: dgibso28@jhu.edu

Abstract

Background: Increased mobile phone penetration allows the interviewing of respondents using interactive voice response surveys in low- and middle-income countries. However, there has been little investigation of the best type of incentive to obtain data from a representative sample in these countries.

Objective: We assessed the effect of different airtime incentives options on cooperation and response rates of an interactive voice response survey in Bangladesh and Uganda.

Methods: The open-label randomized controlled trial had three arms: (1) no incentive (control), (2) promised airtime incentive of 50 Bangladeshi Taka (US $0.60; 1 BDT is approximately equivalent to US $0.012) or 5000 Ugandan Shilling (US $1.35; 1 UGX is approximately equivalent to US $0.00028), and (3) lottery incentive (500 BDT and 100,000 UGX), in which the odds of winning were 1:20. Fully automated random-digit dialing was used to sample eligible participants aged ≥18 years. The risk ratios (RRs) with 95% confidence intervals for primary outcomes of response and cooperation rates were obtained using log-binomial regression.

Results: Between June 14 and July 14, 2017, a total of 546,746 phone calls were made in Bangladesh, with 1165 complete interviews being conducted. Between March 26 and April 22, 2017, a total of 178,572 phone calls were made in Uganda, with 1248 complete interviews being conducted. Cooperation rates were significantly higher for the promised incentive (Bangladesh: 39.3%; RR 1.38, 95% CI 1.24-1.55, *P*<.001; Uganda: 59.9%; RR 1.47, 95% CI 1.33-1.62, *P*<.001) and the lottery incentive arms (Bangladesh: 36.6%; RR 1.28, 95% CI 1.15-1.45, *P*<.001; Uganda: 54.6%; RR 1.34, 95% CI 1.21-1.48, *P*<.001) than those for the control arm (Bangladesh: 28.4%; Uganda: 40.9%). Similarly, response rates were significantly higher for the promised incentive (Bangladesh: 26.5%%; RR 1.26, 95% CI 1.14-1.39, *P*<.001; Uganda: 41.2%; RR 1.27, 95% CI 1.16-1.39, *P*<.001) and lottery incentive arms (Bangladesh: 24.5%%; RR 1.17, 95% CI 1.06-1.29, *P*=.002; Uganda: 37.9%%; RR 1.17, 95% CI 1.06-1.29, *P*=.001) than those for the control arm (Bangladesh: 21.0%; Uganda: 32.4%).

Conclusions: Promised or lottery airtime incentives improved survey participation and facilitated a large sample within a short period in 2 countries.

Trial Registration: ClinicalTrials.gov NCT03773146; http://clinicaltrials.gov/ct2/show/NCT03773146

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e36943) doi:10.2196/36943

KEYWORDS

mobile phone survey; interactive voice response survey; survey; interactive voice response; non-communicable disease; surveillance; airtime incentive; response rate; cooperation rate; communicable disease; Uganda; Bangladesh; low income; middle income; LMIC; Africa; incentive; RCT; randomized controlled trial; lottery; cooperation; participation

Introduction

It is well evidenced that low- and middle-income countries are undergoing demographic and epidemiologic transitions; there is an increasing burden from noncommunicable diseases such as hypertension, diabetes, stroke, and other diseases [1,2]. Four mostly modifiable risk factors primarily contribute to this high noncommunicable disease burden—unhealthy diets, physical inactivity, tobacco use, and excessive alcohol consumption [3,4]. Continuous surveillance and monitoring of these risk factors are crucial to prevent and control noncommunicable diseases [5]. However, collecting data for noncommunicable disease risk factor surveillance is challenging in low- and middle-income countries due to the level of effort, time, and money required for face-to-face interviews and associated data management, analysis, and reporting [6].

High-income countries implement telephone interviews to obtain population-level estimates for health-related outcomes [7,8]. Participation in telephone surveys has declined in recent years in high-income countries, and other survey methods (eg, web-based) are also used to collect health-related data. Although most low- and middle-income countries do not have the infrastructure for conducting landline- or web-based surveys, increased access and ownership of mobile phones in low- and middle-income countries provide the opportunity to use mobile phone numbers for household surveys [9]. Throughout the COVID-19 pandemic, mobile phone surveys have been used to collect data on a broad range of topics [10-13].

There are several options for delivering mobile phone surveys: SMS text messaging, call center interviews by a human operator, and interactive voice response [14]. Interactive voice response is a mobile phone survey method wherein respondents use their mobile phone keypad to select answers from prespecified options (eg, “press 1 if you are male; press 2 if you are female”). Incentives for mobile phone surveys, often delivered as cash, coupons, vouchers, or airtime balances, have been shown to increase survey participation [15]. It could also be considered as compensation for the time spent by participants. Incentives may reduce the amount of time required for data collection by recruiting the optimum number of participants in a shorter time period. In high-income countries, where there is a larger body of evidence on a range of different survey types, providing an incentive beforehand typically produces better survey response than promised or lottery incentives across a [16-18]; however, overall findings have been mixed, and some studies [19,20] show that providing incentives does not improve participation. Past studies [21-24] from low- and middle-income countries have also shown similar mixed results. Studies [21,22] have also shown that delivering incentives to everyone, than using a lottery, can increase participation and reduce cost; there have been limited number of studies [23,24] examining the impact of different incentive amount on the overall survey cost, and investigating these factors would be helpful in understanding the feasibility of mobile phone surveys for future data collection. We aim to fill in these gaps in the literature by assessing the effect of different airtime incentive approaches on the cooperation, response, contact, and refusal rates of an interactive voice response survey for noncommunicable disease behavioral risk factors.

Methods

Study Design

We conducted a randomized controlled trial in Bangladesh (an area of approximately 148,000 km^2 with an estimated population of 160 million [25]) and Uganda (an area of approximately 241,000 km^2 with an estimated population of 40 million [25]). In 2017, mobile phone subscription rates were 83 and 55 subscribers per 100 people in Bangladesh and Uganda, respectively [9].

In this trial, incentives were delivered as airtime (ie, mobile phone balance). Participants were randomized to 1 of 3 study arms: no incentive (control arm), a promised airtime incentive of 50 Bangladeshi Taka (US $0.60; 1 BDT is approximately equivalent to US $0.012) or 5000 Ugandan Shilling (US $1.35; 1 UGX is approximately equivalent to US $0.00028) for completing the interactive voice response survey, or lottery incentive (500 BDT and 100,000 UGX), wherein the odds of winning were 1:20. The conduct, analysis, and reporting of results were performed in accordance with Consolidated Standards of Reporting Trials guidelines [26].

Participants

Participants were sampled using a fully automated random-digit dialing method [27]. Briefly, the country codes along with the 3-digit sequence specific to the mobile network operator were used as the base for potential mobile phone numbers. The remaining 7 digits were generated randomly. Respondents who self-reported being aged 18 years or older were eligible for the trial. The survey was programmed to have a designated local number appear on the respondent’s caller ID screen.

Randomization and Masking

The automated randomization process was performed within the interactive voice response platform to cover all mobile phone networks in both countries. Participants were randomized after selecting the survey language but prior to consent (Figure S1 in Multimedia Appendix 1). Due to the nature of the study design, participants were informed about the incentive during the survey introduction and, therefore, were not blinded to study arm allocation. Statisticians involved in data cleaning and analysis were blinded to participant allocation.

XSL•FO
RenderX

Procedures

The overall procedures were similar in both countries. Interactive voice response surveys were sent only once to each phone number, and calls were made between 8 AM and 8 PM local time. Respondents who picked up the phone were instructed to select a language from a list of languages: Bangla or English in Bangladesh and Luganda, Luo, Runyakitara, or English in Uganda. Candidate participants listened to a description of the survey objectives and requirements for the incentive (ie, survey completion) as applicable (Table S1 and Figure S1 in Multimedia Appendix 1). Participants were told that they would not incur any expenses by answering the survey. Age-eligibility was confirmed (ie, "Are you 18 years or older? If yes, press 1; if no, press 3"). Age-eligible candidates were provided the consent disclosure statement and asked to authorize their participation by pressing the 1 button on the mobile phone. Participants answered demographic and noncommunicable disease questions, and only those who completed the survey received the incentive. Participants were instructed to press the star key to repeat any questions.

Demographic data on age, gender, education, and location were collected to perform subgroup analysis (ie, to identify differences in participation by those characteristics). Noncommunicable disease questions were grouped into 5 modules: tobacco use, alcohol consumption, dietary habits (including consumption of fruits, vegetables, and salt), physical activity, and medical conditions (including hypertension and diabetes). Because respondents could end the interview before finishing all modules, the order of the noncommunicable disease modules was randomized to minimize attrition and to ensure that data reporting errors were as randomly distributed as possible. Questions within a module were not randomized in order to maintain skip patterns. The questionnaire was adapted from standardized surveys [28], and initial cognitive testing and user groups were conducted with people who identified themselves as being from a low- and middle-income country at Johns Hopkins University [29]. A series of key informant interviews and focus group discussions were also conducted in each country to assess the comprehensibility and accuracy of translated questionnaires and to improve the usability of the interactive voice response platform.

Ethical Approval

Johns Hopkins Bloomberg School of Public Health, Makerere University School of Public Health, The Uganda National Council for Science and Technology, and The Institute of Epidemiology Disease Control and Research institutional review boards approved the study protocol (number NCT03773146). The study was registered (NCT03773146), and the study protocol has been published elsewhere [30].

Outcomes

The primary outcomes of this trial were response rates 4 and cooperation rates 1, as defined by the American Association for Public Opinion Research (Table S2 in Multimedia Appendix 1) [31]. Response rate calculations included partial and complete surveys in the numerator. Cooperation rate was calculated as the proportion of complete interviews from all eligible respondents, but the calculation did not include people who immediately hung up or who did not answer the age question in the denominator. Secondary outcomes were contact rate 2 and refusal rate 2 [31]. The cooperation rate was the number of complete interviews divided by the sum of complete, partial, and noninterviews. Complete interviews were defined as respondents who answered at least 4 of the 5 noncommunicable disease modules. Partial interviews were defined as respondents who answered between 1 and 3 noncommunicable disease modules. Noninterviews included refusals (ie, participants who ended the survey at the consent question) and break-offs (ie, participants who were 18 years or older but did not complete an noncommunicable disease module). The response rate was calculated as the number of complete and partial interviews divided by the total number of complete and partial interviews, refusals, break-offs, and the estimated proportion of age-eligibility unknown calls (individuals who initiated the survey but did not answer the age question). The estimated proportion of unknown eligibility was obtained from the proportion of participants who responded to the age-screening question and indicated they were 18 years or older. Calls were classified as ineligible if the individual indicated an age below 18 years or did not pick up the phone. As a secondary analysis, contact refusal and response rates were calculated without applying e for the unknown participants.

Statistical Analysis

Demographic characteristics of complete interviews were described by study arms and compared using chi-square tests. Using the control arm as the reference category, risk ratios (RR) and 95% confidence intervals were calculated for contact, response, refusal, and cooperation rates with log-binomial regression [32]. To assess any potential effect modification of incentives on cooperation rates, the log-binomial models were extended and interaction terms with education, gender, age, and location were tested. We did not assess any effect modification for response rates because its equation included disposition codes for *Unknown* (participants who did not answer any of the demographic questions).

We calculated pooled risk ratios for different incentive arms using random-effects meta-analysis [33]. The heterogeneity statistic (ie, I^2) was estimated using the Mantel-Haenszel method. The I^2 statistic indicates the proportion of variability in effect that resulted from heterogeneity instead of chance or sampling error. A lower I^2 statistic suggests lower heterogenicity. We calculated the direct delivery cost per complete survey, which included the cost of airtime used to deliver the survey and the incentive amount, as applicable. We summed the total call durations by arm and multiplied by per-minute airtime cost (US $0.04 in Bangladesh and $0.10 in Uganda) to produce the estimated cost per completed survey. Stata (version 14.0; StataCorp LLC) was used to analyze data. An α=.05 was assumed for all tests of statistical significance.

Sample Size

We used the same assumptions to calculate required sample sizes for the trial in both countries. With a 30% cooperation rate of the control arm, 5% type 1 error, and 80% power, 376

XSL•FO
RenderX

participants were required to complete the interview for each study arm in order to detect a 10% difference between control and incentive arms. The total required sample size (ie, complete surveys) was 1128 in each country. As recommended [34], we did not inflate the sample for multiple comparisons.

Results

From June 14, 2017 to July 14, 2017, a total of 1165 compete interviews were obtained from 546,746 phone calls in Bangladesh (Figure 1). In Uganda, 178,572 calls were made between March 26 and April 22, 2017 to obtain 1248 complete interviews (Figure 2). In both countries, the sociodemographic characteristics of complete interviews were similar across study arms (Table 1). Of 1165 respondents in Bangladesh, 89.4% (n=1042) respondents were male. Of 1248 respondents in Uganda, 76.0% (n=948) respondents were male. Most respondents were between the ages of 18 to 29 years old—74.4% (867/1165) and 71.0% (886/1248) in Bangladesh and Uganda, respectively.

Figure 1. Consolidated Standard of Reporting Trial diagram of study participants in Bangladesh.

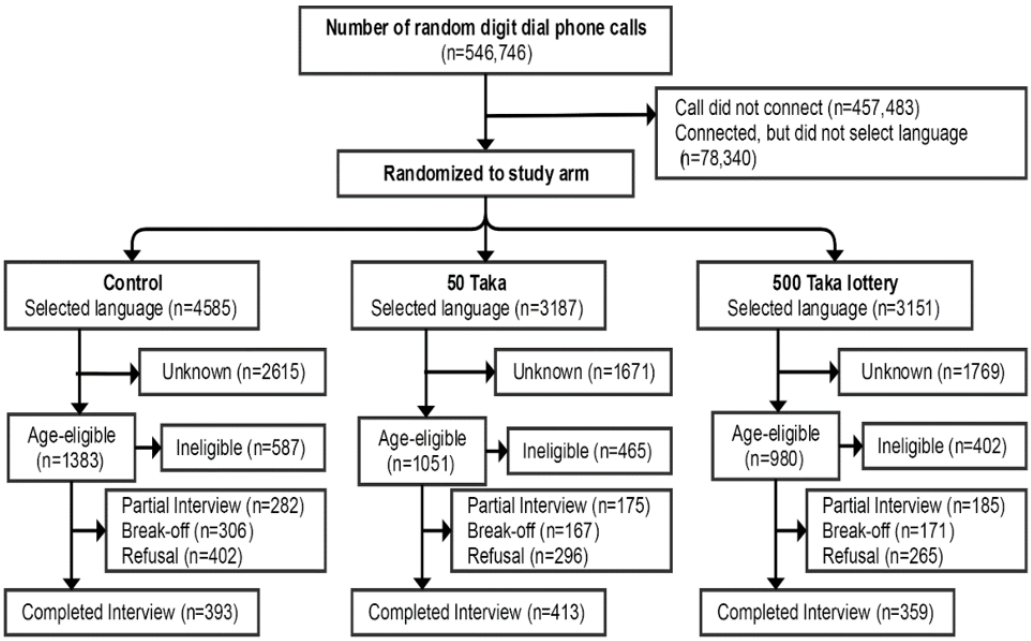
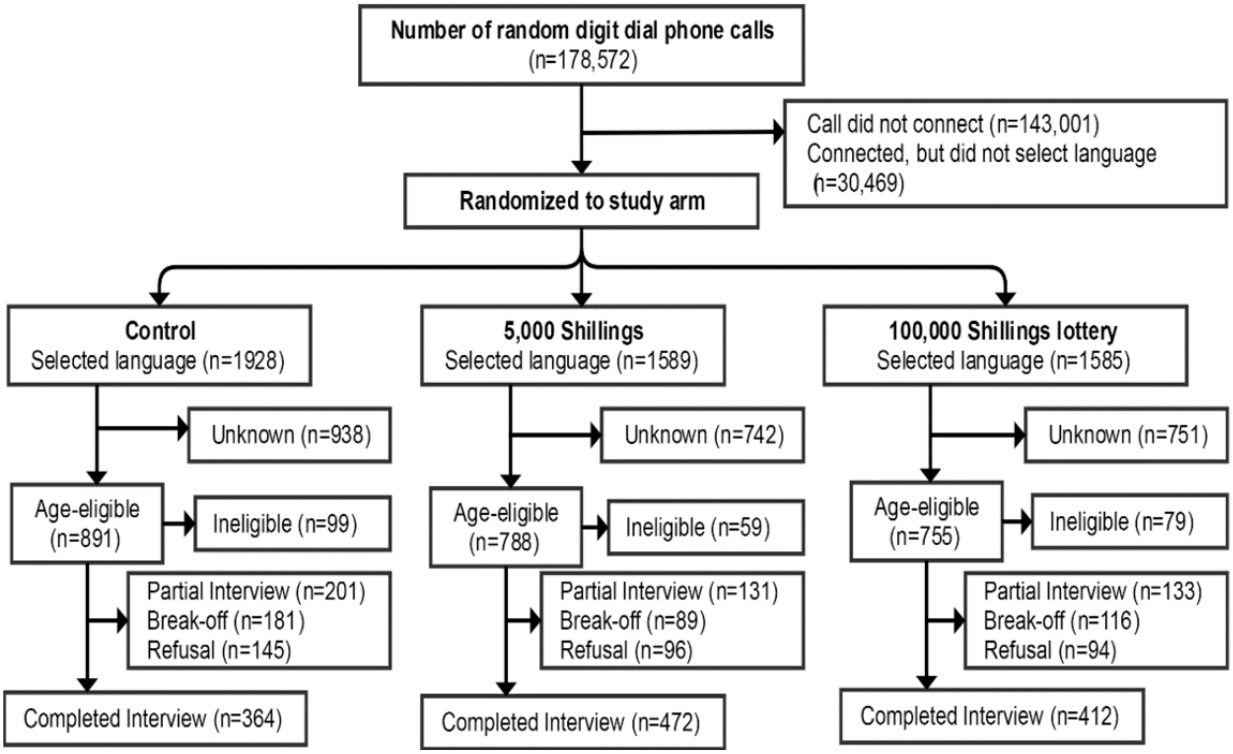


Figure 2. Consolidated Standard of Reporting Trial diagram of study participants in Uganda.



XSL•FO
RenderX

Table 1. Demographic characteristics of complete interviews by study arm.

	Bangladesh (n=1165), n (%) or n			Uganda (n=1248), n (%) or n		
	Control (n=393)	Promised incentive (n=413)	Lottery incentive (n=359)	Control (n=364)	Promised incentive (n=472)	Lottery incentive (n=412)
Sex						
Male	353 (89.8)	369 (89.4)	320 (89.1)	276 (77.5)	357 (77.6)	315 (78.7)
Female	39 (9.9)	44 (10.7)	37 (10.3)	80 (22.5)	103 (22.4)	85 (21.3)
Other	1 (0.3)	0 (0.0)	2 (0.6)	N/A[a]	N/A	N/A
Missing	0	0	0	8	12	12
Age group (years)						
18-29	301 (76.6)	291 (70.5)	275 (76.6)	257 (72.2)	326 (71.0)	303 (75.6)
30-49	75 (19.1)	109 (26.4)	71 (19.8)	91 (25.6)	124 (27.0)	92 (23.0)
50-69	10 (2.5)	9 (2.2)	9 (2.5)	6 (1.7)	7 (1.5)	3 (0.8)
70+	7 (1.8)	4 (1.0)	4 (1.1)	2 (0.6)	2 (0.4)	3 (0.8)
Missing	0	0	0	8	13	11
Education attempted						
None	29 (22.3)	34 (25.4)	28 (20.7)	65 (18.3)	63 (13.5)	59 (14.5)
Primary	100 (76.9)	100 (74.6)	107 (79.3)	83 (23.4)	114 (24.4)	107 (26.3)
Secondary	N/A	N/A	N/A	146 (41.1)	209 (44.8)	169 (41.5)
Tertiary or higher	N/A	N/A	N/A	61 (17.2)	81 (17.3)	72 (17.7)
Refused	1 (0.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	263	279	224	9	5	5
Location						
Urban	225 (57.2)	222 (53.8)	180 (50.1)	178 (49.7)	250 (54.1)	227 (56.3)
Rural	165 (42.0)	191 (46.2)	179 (49.9)	180 (50.3)	212 (45.9)	176 (43.7)
Refused	3 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	0	0	0	6	10	9
Language						
Bangla	390 (99.2)	410 (99.3)	355 (98.9)	N/A	N/A	N/A
English	3 (0.8)	3 (0.7)	4 (1.1)	56 (15.4)	68 (14.4)	66 (16.0)
Luganda	N/A	N/A	N/A	213 (58.5)	260 (55.2)	248 (60.2)
Luo	N/A	N/A	N/A	36 (9.9)	50 (10.6)	29 (7.0)
Runyakitara	N/A	N/A	N/A	59 (16.2)	93 (19.8)	69 (16.8)
Missing	0	0	0	0	1	0

[a]N/A: not applicable.

The sociodemographic characteristics of respondents with complete and partial interviews were similar in both countries, with the exception of a significant difference in age distribution in Bangladesh (*P*=.002); complete interviews had higher proportion of respondents aged 18 to 29 years old than partial interviews (Table S3 in Multimedia Appendix 1). The median time spent completing the interactive voice response survey was 15 minutes 8 seconds (IQR 14 minutes 8 seconds to 16 minutes 15 seconds) and 13 minutes 38 seconds (IQR 12 minutes 39 seconds to 14 minutes 45 seconds) in Bangladesh and Uganda, respectively. The direct costs of airtime, and incentives where applicable per complete interview were $3.89 and $3.16 in the control arm, $3.90 and $3.91 in the promised incentive arm, and $4.05 and $4.12 in the lottery incentive arm, in Bangladesh and Uganda, respectively (Table 2).

Cooperation and response rates were significantly higher in the promised incentive arm (cooperation: 413/1051, 39.3%; RR 1.38, 1.24-1.55, *P*<.001; response: 588/2222, 26.5%, RR 1.26, 95% CI 1.14-1.39, *P*<.001) and in the lottery arm (cooperation: 359/980, 36.6%; RR 1.28, 95% CI 1.15-1.45, *P*<.001; response: 544/2220, 24.5%; RR 1.17, 95% CI 1.06-1.29, *P*=.002) compared with those for the control arm (cooperation: 393/1383, 28.4%; response: 675/3216, 21.0%). In Uganda, the cooperation

XSL•FO
RenderX

and response rates were higher than those in Bangladesh. Rates were significantly higher in the promised (cooperation: RR 1.47, 95% CI 1.33-1.62, P<.001; response: RR 1.27, 95% CI 1.16-1.39, P<.001) and lottery arms (cooperation: RR 1.34, 95% CI 1.21-1.48, P<.001; response: RR 1.17, 95% CI 1.06-1.29, P=.001) compared with those for the control arm. In both countries, cooperation and response rates were similar when using equations that did not include the estimated proportion of age-eligible participants in the unknown disposition code (Table S4 in Multimedia Appendix 1). In both countries, subgroup analyses showed that participant gender, age, education, and location did not modify the intervention's effect on cooperation rate (Tables S5 and S6 in Multimedia Appendix 1).

Pooling Bangladesh and Uganda participants showed that the promised incentive (pooled RR 1.42, 95% CI 1.32-1.53, P<0.001) and lottery incentive (pooled RR 1.31, 95% CI 1.21-1.41, P<0.001) significantly improved cooperation rate compared with no incentive (Figure 3). Similarly, response rates were significantly higher in the promised incentive (pooled RR 1.26, 95% CI 1.18-1.35, P<0.001) and lottery incentive (pooled RR 1.17, 95% CI 1.09-1.25, P<0.001,I^2=0.0%) arm compared with that in the control arm. Overall, any incentive significantly improved cooperation rates by 37% (pooled RR 1.37, 95% CI 1.29-1.44, P<0.001) and response rates by 22% (pooled RR 1.22, 95% CI 1.18-1.28, P<.001), and these results were highly consistent (cooperation: I^2=12.1%, P=.33; response: I^2=0.0%, P=.47).

Table 2. Disposition codes and survey rates by study arm.

	Bangladesh			Uganda		
	Control	Promised incentive	Lottery incentive	Control	Promised incentive	Lottery incentive
Complete interview, n	393	413	359	364	472	412
Partial interview, n	282	175	185	201	131	133
Refusal						
Refusal	402	296	265	145	96	94
Breaks-off	306	167	171	181	89	116
Unknown other, n	2615	1671	1769	938	742	751
Estimated unknown[a]	1833	1171	1240	854	675	684
Ineligible, n						
Under age	587	465	402	99	59	79
Call did not connect[b]	152,494	152,494	152,495	47,667	47,667	47,667
Connected, but no language selection[b]	26,114	26,113	26,113	10,156	10,156	10,157
Average cost (US $) per complete interview[c]	3.89	3.90	4.05	3.16	3.91	4.12
Contact rate	43.00	47.30	44.10	51.10	53.90	52.50
Risk ratio (95% CI)	Ref	1.10 (1.04-1.17)	1.03 (0.97-1.09)	Ref	1.05 (0.99-1.13)	1.03 (0.96-1.10)
P value	Ref	.002	0.40	Ref	.11	.43
Response rate	21.00	26.50	24.50	32.40	41.20	37.90
Risk ratio (95% CI)	Ref	1.26 (1.14-1.39)	1.17 (1.06-1.29)	Ref	1.27 (1.16-1.39)	1.17 (1.06-1.29)
P value	Ref	<.001	.002	Ref	<.001	.001
Refusal rate	22.00	20.80	19.60	18.70	12.70	14.60
Risk ratio (95% CI)	Ref	0.95 (0.85- 1.05)	0.89 (0.80- 0.99)	Ref	0.68 (0.57- 0.80)	0.78 (0.67- 0.92)
P value	Ref	.30	.04	Ref	<.001	.002
Cooperation rate	28.40	39.30	36.60	40.90	59.90	54.60
Risk ratio (95% CI)	Ref	1.38 (1.24-1.55)	1.28 (1.15-1.45)	Ref	1.47 (1.33-1.62)	1.34 (1.21-1.48)
P value	Ref	<.001	<.001	Ref	<.001	<.001

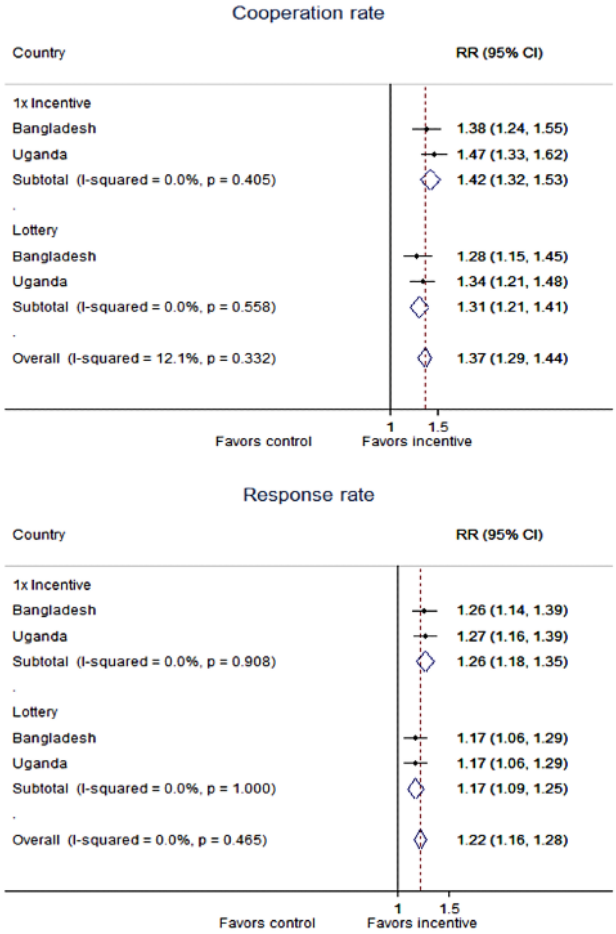
[a]Estimated proportion of unknown cases that were age-eligible was 70.1% for Bangladesh and 91.0% for Uganda.

[b]Evenly distributed to each study arm due to randomization occurring after language selection.

[c]Only includes cost of the call based on time participants spent on the survey plus airtime incentive, as applicable.

XSL•FO
RenderX

Figure 3. Pooled risk ratios for cooperation and response rate by study arm.



Discussion

In this study, the promised and lottery incentive arms had higher response and cooperation rates than control arms (ie, no incentive arm) in both Bangladesh and Uganda. Given that the pooled analysis showed that the incentives were highly consistent at increasing these survey rates in two culturally and geographically distinct low- and middle-income countries, the provision of airtime incentives may be a useful mechanism to increase interactive voice response survey participation in other low-resource settings.

There are very few studies [4,35] that have examined the use of airtime incentives in low- and middle-income countries. Our study adds significant knowledge to the growing body of literature on the impact of incentive amount on mobile phone survey in these countries. A previous study [4] similarly found that both promised and lottery airtime incentives significantly improved the completion rate of a random-digit dialing interactive voice response survey in Zimbabwe. In Mozambique, the lottery incentive, but not the promised incentive, increased completion rates [35].

A study [36] from Honduras also found that providing either US $1 or $5 of airtime significantly improved response rates compared with the no incentive arm. Another study [22] showed that providing an airtime incentive of at least 50 BDT in Bangladesh and 5000 UGX in Uganda improved survey participation compared to people without any incentive and also reduced the number of incomplete interviews.

Our interactive voice response survey data collection was quick and inexpensive. In Bangladesh, we collected 1165 complete interviews in 21 days at a cost of approximately US $4.00 per complete interview. In Uganda, 1248 complete interviews were collected in 18 days at a cost under US $4.00 per complete interview. Our findings are similar to those from a random-digit dialing interactive voice response survey in Ghana collected 9469 complete interviews in 27 days at a cost of US $4.95 per complete interview [37]. The average cost of a competed interview is much lower than the average cost of such a household survey, this indicates that the mobile phone survey could be cost-effective compared to household surveys. For instance, Lietz and colleagues [38] estimated the average cost of per completed interview of the Nouna Health and Demographic Survey in rural Burkina Faso as approximately US $25. Although the specific objectives of that survey were broader and required a longer amount of time than our survey, future studies should compare the average cost of conducting an interview in such mobile phone surveys. In Bangladesh, our use of an airtime incentive to motivate participants to complete the interview became cost-neutral compared to the control arm. The savings in cost was due to the decreased number of partial interviews and, therefore, fewer phone calls. We did not see a similar finding in Uganda where the promised (US $1.35) and lottery (US $28) incentive amounts were higher than those in Bangladesh (promised: US $0.60; lottery: US $6.00). The

XSL•FO
RenderX

difference in promised incentive amount may also account for some differences in participation rate by country. Specifically, people may not initiate a survey if the promised incentive amount appears low, which would ultimately reduce participation. Future work could manipulate the odds of winning the lottery and its amount to ensure the incentive is cost-neutral or even cost-saving [39].

Our cooperation and response rates were calculated in a standardized manner using American Association for Public Opinion Research guidelines [31], which allows for comparison with other studies. In a nationally administered random-digit dialing interactive voice response survey, with persons ≥18 years in Ghana, in which no incentives were provided, contact (39%) and response (31%) rates were similar to those observed in our control arms for Bangladesh and Uganda [37]. However, we observed higher refusal and lower cooperation rates in Bangladesh (refusal: 22%; cooperation: 28%) and Uganda (refusal: 19%; cooperation: 41%) than what was observed in Ghana (refusal: 7%; cooperation: 59%). These differences may be explained by variations in the eligibility criteria, length of survey, and the classification of disposition codes for complete, break-offs, refusals, and partial interviews. For instance, L'Engle and colleagues [37] defined complete interviews as responding to all survey questions, while we defined complete as 4 out of 5 modules.

There are a range of ethical considerations in mobile phone survey [40]. Our survey started with an introduction that included the purpose of the study, the sponsoring agency, time commitment, and that the data would be kept confidential. Participants were offered an opportunity to consent to the survey by pressing a button on their mobile phone and were allowed to refuse to answer any question. Additional studies that evaluate alternative ways to consent participants are needed to maximize participant's understanding of the study [41]. Additionally, there has been considerable discussion on the ethics of incentives and health research [42,43]. Our use of incentives was informed by in-country stakeholders, amounts used were less than a day's working wage and were not paired with risky or unsafe behavior. Nonetheless, we believe important to acknowledge that efforts to optimize use of incentives, in general, should be informed not only by cost-effectiveness considerations. Incentives that insufficiently reflect response burden, or that, perhaps in rare cases, have the potential to unduly influence or induce participation, ought to be avoided.

We observed a higher proportion of male, young (ie, 18 to 29 years old), or urban residents compared to general population in both countries. This finding was similar to those of random-digit dialing interactive voice response surveys conducted in Afghanistan, Ethiopia, Ghana, Mozambique, Tanzania, and Zimbabwe [35,37,44]. Male gender, younger age, higher education, and urban residence have been found to be associated with mobile phone ownership in low- and middle-income countries, including East Africa [45] and Bangladesh [46]. This does raise concerns about the ability to generate nationally representative estimates (ie, generalizability of the findings). Advances in sampling and statistical methodology may be required for such estimates. Quota sampling could be used to ensure a more equal distribution of the sociodemographic characteristics [47]. Others have found that weighted estimates of noncommunicable disease indicators collected via mobile phone survey approximate household collected data [48].

This study has several strengths. First, the randomization was automated and embedded within the interactive voice response platform. This safeguarded against misallocation of participants to study arm which could bias response and cooperation rates. Second, we employed standardized protocols and questionnaires in both countries and used the same technology platform to deliver interactive voice response surveys to afford for cross-country comparisons. Lastly, our sampling frame consisted of all known mobile network operators in each country; thereby minimizing potential selection bias.

In addition to underrepresentation from some sociodemographic populations, this study has some limitations. First, there was a substantial number of phone calls in Bangladesh and Uganda where we were unable to determine the status of the phone numbers. Calling people randomly can also reduce response. We could not determine if the phone numbers we called were active or inactive numbers [49]. As randomization to study arm occurred after participants picked up the phone, we chose to designate these phone calls as nonworking numbers. This decision inflates our contact, response, and refusal rates, but has no effect on the cooperation rate. Second, although not an issue in Bangladesh where 99% of the respondents took the interactive voice response survey in Bangla, our survey was only available in 3 of the 6 major language groups in Uganda [50]. This might lead to some selection error due to unavailability of the preferred language and would have larger implications for nationally representative surveys [35]. We did not check the quality of collected data as that was not the main purpose of this study; future studies should investigate that.

We investigated the response, contact, and cooperation rates of 2 different incentive structures compared to providing no incentives in 2 geographically and linguistically, distinct countries. We observed that providing either type of incentive enhanced survey participation and minimized associated costs.

Acknowledgments

This study was made possible by the generous support of Bloomberg Philanthropies (grant 41388.01) to AAH and the people of Australia through the Department of Foreign Affairs and Trade. The contents are the responsibility of the authors and do not necessarily reflect the views of Bloomberg Philanthropies or the Government of Australia. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

XSL•FO
RenderX

We are thankful to our collaborators and research teams at the Institute of Epidemiology, Disease Control and Research, Bangladesh and Makerere University School of Public Health, Uganda. We are also thankful to Viamo (formerly VOTOmobile) and colleagues Kevin Schuster and Tom Mortimore for their assistance with the interactive voice response platform. Our thanks also go to our research collaborators in the noncommunicable disease component at the Centers for Disease Control and Prevention, the Centers for Disease Control and Prevention Foundation, the World Health Organization, and all those who played a part in developing initial versions of the questionnaires used; who participated in translating survey instruments, collecting data during the formative phase, and making audiorecordings for interactive voice response; or who contributed in any way in both countries.

Data Availability

All data files are available from the OpenICPSR repository database (accession number: ICPSR 107284).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supporting information.
[DOCX File , 64 KB - jmir_v24i5e36943_app1.docx]

Multimedia Appendix 2
CONSORT eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1354 KB - jmir_v24i5e36943_app2.pdf]

References

1. Boutayeb A. The double burden of communicable and non-communicable diseases in developing countries. Trans R Soc Trop Med Hyg 2006 Mar;100(3):191-199. [doi: 10.1016/j.trstmh.2005.07.021] [Medline: 16274715]
2. Bygbjerg IC. Double burden of noncommunicable and infectious diseases in developing countries. Science 2012 Sep 21;337(6101):1499-1501. [doi: 10.1126/science.1223466] [Medline: 22997329]
3. Global status report on noncommunicable diseases 2010. World Health Organization Institutional Repository for Information Sharing. URL: https://apps.who.int/iris/bitstream/handle/10665/148114/9789241564854_eng.pdf [accessed 2022-04-19]
4. Global Burden of Disease Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Br Dent J 2015 Oct 9;219(7):329-329. [doi: 10.1038/sj.bdj.2015.751]
5. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. World Health Organization. 2013. URL: https://www.who.int/publications/i/item/9789241506236 [accessed 2018-08-24]
6. Calba C, Goutard FL, Hoinville L, Hendrikx P, Lindberg A, Saegerman C, et al. Surveillance systems evaluation: a systematic review of the existing approaches. BMC Public Health 2015;15:448 [FREE Full text] [doi: 10.1186/s12889-015-1791-5] [Medline: 25928645]
7. Pickens CM, Pierannunzi C, Garvin W, Town M. Surveillance for certain health behaviors and conditions among states and selected local areas - behavioral risk factor surveillance system, United States, 2015. MMWR Surveill Summ 2018 Jun 29;67(9):1-90 [FREE Full text] [doi: 10.15585/mmwr.ss6709a1] [Medline: 29953431]
8. Healy J, Rodriguez-Lainz A, Elam-Evans LD, Hill HA, Reagan-Steiner S, Yankey D. Vaccination coverage among foreign-born and U.S.-born adolescents in the United States: successes and gaps - National Immunization Survey-Teen, 2012-2014. Vaccine 2018 Mar 20;36(13):1743-1750 [FREE Full text] [doi: 10.1016/j.vaccine.2018.02.052] [Medline: 29483032]
9. Measuring the information society report 2017. International Telecommunication Union. 2017. URL: https://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2017.aspx [accessed 2022-01-01]
10. Byonanebye DM, Nabaggala MS, Naggirinya AB, Lamorde M, Oseku E, King R, et al. An interactive voice response software to improve the quality of life of people living with HIV in Uganda: randomized controlled trial. JMIR Mhealth Uhealth 2021 Feb 11;9(2):e22229 [FREE Full text] [doi: 10.2196/22229] [Medline: 33570497]
11. Akhtar Z, Mah-E-Muneer S, Rashid MM, Ahmed MS, Islam MA, Chowdhury S, et al. Antibiotics use and its knowledge in the community: a mobile phone survey during the COVID-19 pandemic in Bangladesh. Antibiotics (Basel) 2021 Aug 29;10(9):1052 [FREE Full text] [doi: 10.3390/antibiotics10091052] [Medline: 34572634]
12. Lamichhane A, Rana S, Shrestha K, Paudyal R, Malla P, Upadhyaya S, et al. Violence and sexual and reproductive health service disruption among girls and young women during COVID-19 pandemic in Nepal: a cross-sectional study using interactive voice response survey. PLoS One 2021 Dec 8;16(12):e0260435 [FREE Full text] [doi: 10.1371/journal.pone.0260435] [Medline: 34879111]

XSL•FO
RenderX

13. Zheng Y, Wang W, Zhong Y, Wu F, Zhu Z, Tham Y, et al. A peer-to-peer live-streaming intervention for children during COVID-19 homeschooling to promote physical activity and reduce anxiety and eye strain: cluster randomized controlled trial. J Med Internet Res 2021 Apr 30;23(4):e24316 [FREE Full text] [doi: 10.2196/24316] [Medline: 33882021]
14. Hyder AA, Wosu AC, Gibson DG, Labrique AB, Ali J, Pariyo GW. Noncommunicable disease risk factors and mobile phones: a proposed research agenda. J Med Internet Res 2017 May 05;19(5):e133 [FREE Full text] [doi: 10.2196/jmir.7246] [Medline: 28476722]
15. Greenleaf AR, Gibson DG, Khattar C, Labrique AB, Pariyo GW. Building the evidence base for remote data collection in low- and middle-income countries: comparing reliability and accuracy across survey modalities. J Med Internet Res 2017 May 05;19(5):e140 [FREE Full text] [doi: 10.2196/jmir.7331] [Medline: 28476728]
16. Singer E, Ye C. The use and effects of incentives in surveys. Ann Am Acad Pol Soc Sci 2012 Nov 26;645(1):112-141. [doi: 10.1177/0002716212458082]
17. Brick J, Montaquila J, Hagedorn M, Roth S, Chapman C. Implications for RDD design from an incentive experiment. J Off Stat 2005;21(4):571-589 [FREE Full text]
18. Smith MG, Witte M, Rocha S, Basner M. Effectiveness of incentives and follow-up on increasing survey response rates and participation in field studies. BMC Med Res Methodol 2019 Dec 05;19(1):230 [FREE Full text] [doi: 10.1186/s12874-019-0868-8] [Medline: 31805869]
19. Gjøstein DK, Huitfeldt A, Løberg M, Adami H, Garborg K, Kalager M, et al. Incentives and participation in a medical survey. Tidsskr Nor Laegeforen 2016 Jul;136(12-13):1082-1087 [FREE Full text] [doi: 10.4045/tidsskr.15.1096] [Medline: 27381786]
20. Pieper D, Kotte N, Ober P. The effect of a voucher incentive on a survey response rate in the clinical setting: a quasi-randomized controlled trial. BMC Med Res Methodol 2018 Aug 16;18(1):86 [FREE Full text] [doi: 10.1186/s12874-018-0544-4] [Medline: 30115037]
21. Chamie G, Ndyabakira A, Marson KG, Emperador DM, Kamya MR, Havlir DV, et al. A pilot randomized trial of incentive strategies to promote HIV retesting in rural Uganda. PLoS One 2020 May 29;15(5):e0233600 [FREE Full text] [doi: 10.1371/journal.pone.0233600] [Medline: 32470089]
22. Gibson DG, Wosu AC, Pariyo GW, Ahmed S, Ali J, Labrique AB, et al. Effect of airtime incentives on response and cooperation rates in non-communicable disease interactive voice response surveys: randomised controlled trials in Bangladesh and Uganda. BMJ Glob Health 2019;4(5):e001604 [FREE Full text] [doi: 10.1136/bmjgh-2019-001604] [Medline: 31565406]
23. Göritz A. Incentives in web studies: methodological issues and a review. Int J Internet Sci 2006;1(1):58-70 [FREE Full text]
24. Gibson DG, Pereira A, Farrenkopf BA, Labrique AB, Pariyo GW, Hyder AA. Mobile phone surveys for collecting population-level estimates in low- and middle-income countries: a literature review. J Med Internet Res 2017 May 05;19(5):e139 [FREE Full text] [doi: 10.2196/jmir.7428] [Medline: 28476725]
25. The world factbook. Central Intelligence Agency. 2019. URL: https://www.cia.gov/the-world-factbook/ [accessed 2019-07-22]
26. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. BMC Med 2010 Mar 24;8:18 [FREE Full text] [doi: 10.1186/1741-7015-8-18] [Medline: 20334633]
27. Waksberg J. Sampling methods for random digit dialing. J Am Stat Assoc 1978 Mar;73(361):40-46. [doi: 10.1080/01621459.1978.10479995]
28. The STEPS instrument and support materials. World Health Organization. 2018. URL: https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps [accessed 2018-08-24]
29. Gibson DG, Farrenkopf BA, Pereira A, Labrique AB, Pariyo GW. The development of an interactive voice response survey for noncommunicable disease risk factor estimation: technical assessment and cognitive testing. J Med Internet Res 2017 May 05;19(5):e112 [FREE Full text] [doi: 10.2196/jmir.7340] [Medline: 28476724]
30. Gibson DG, Pariyo GW, Wosu AC, Greenleaf AR, Ali J, Ahmed S, et al. Evaluation of mechanisms to improve performance of mobile phone surveys in low- and middle-income countries: research protocol. JMIR Res Protoc 2017 May 05;6(5):e81 [FREE Full text] [doi: 10.2196/resprot.7534] [Medline: 28476729]
31. The American Association for Public Opinion Research. Standard definitions: final dispositions of case codes and outcome rates for surveys. Encyclopedia Survey Research Methods. 2016. URL: https://www.aapor.org/aapor_main/media/publications/standard-definitions20169theditionfinal.pdf [accessed 2018-01-01]
32. Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. Am J Epidemiol 1986 Jan;123(1):174-184. [doi: 10.1093/oxfordjournals.aje.a114212] [Medline: 3509965]
33. Deeks J, Altman D, Bradburn M. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Ingger M, Smith G, Altman D, editors. Systematic Reviews in Health Care: Meta - Analysis in Context, Second edition. London, United Kingdom: BMJ Publishing Group; 2008:312.
34. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology 1990;1(1):43-46. [doi: 10.1097/00001648-199001000-00010] [Medline: 2081237]
35. Leo B, Morello R, Mellon J, Piexoto T, Davenport S. Do mobile surveys work in poor countries. International Household Survey Network. 2012. URL: https://www.cgdev.org/sites/default/files/CGD-Working-Paper-398-Mobile-Phones.pdf [accessed 2018-10-05]

XSL•FO
RenderX

36. Ballivian A, Azevedo J, Durbin W. Using mobile phones for high frequency data collection. In: Toninelli D, Pinter R, de Pedraza P, editors. Mobile Research Methods: Opportunities and Challenges of Mobile Research Methodologies. London: Ubiquity Press; 2013.
37. L'Engle K, Sefa E, Adimazoya EA, Yartey E, Lenzi R, Tarpo C, et al. Survey research with a random digit dial national mobile phone sample in Ghana: methods and sample quality. PLoS One 2018;13(1):e0190902 [FREE Full text] [doi: 10.1371/journal.pone.0190902] [Medline: 29351349]
38. Lietz H, Lingani M, Sié A, Sauerborn R, Souares A, Tozan Y. Measuring population health: costs of alternative survey approaches in the Nouna Health and Demographic Surveillance System in rural Burkina Faso. Glob Health Action 2015 Aug 07;8(1):28330 [FREE Full text] [doi: 10.3402/gha.v8.28330] [Medline: 26257048]
39. Gajic A, Cameron D, Hurley J. The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. Eur J Health Econ 2012 Dec;13(6):789-799. [doi: 10.1007/s10198-011-0332-0] [Medline: 21691841]
40. Ali J, Labrique AB, Gionfriddo K, Pariyo G, Gibson DG, Pratt B, et al. Ethics considerations in global mobile phone-based surveys of noncommunicable diseases: a conceptual exploration. J Med Internet Res 2017 May 05;19(5):e110 [FREE Full text] [doi: 10.2196/jmir.7326] [Medline: 28476723]
41. Ali J, DiStefano MJ, Coates McCall I, Gibson DG, Al Kibria GM, Pariyo GW, et al. Ethics of mobile phone surveys to monitor non-communicable disease risk factors in low- and middle-income countries: a global stakeholder survey. Glob Public Health 2019 Aug 10;14(8):1167-1181. [doi: 10.1080/17441692.2019.1566482] [Medline: 30628548]
42. Mduluza T, Midzi N, Duruza D, Ndebele P. Study participants incentives, compensation and reimbursement in resource-constrained settings. BMC Med Ethics 2013;14 Suppl 1:S4 [FREE Full text] [doi: 10.1186/1472-6939-14-S1-S4] [Medline: 24564948]
43. Lunze K, Paasche-Orlow MK. Financial incentives for healthy behavior: ethical safeguards for behavioral economics. Am J Prev Med 2013 Jun;44(6):659-665. [doi: 10.1016/j.amepre.2013.01.035] [Medline: 23683984]
44. Pariyo GW, Greenleaf AR, Gibson DG, Ali J, Selig H, Labrique AB, et al. Does mobile phone survey method matter? reliability of computer-assisted telephone interviews and interactive voice response non-communicable diseases risk factor surveys in low and middle income countries. PLoS One 2019 Apr 10;14(4):e0214450 [FREE Full text] [doi: 10.1371/journal.pone.0214450] [Medline: 30969975]
45. Khatun F, Heywood AE, Hanifi SMA, Rahman MS, Ray PK, Liaw S, et al. Gender differentials in readiness and use of mHealth services in a rural area of Bangladesh. BMC Health Serv Res 2017 Aug 18;17(1):573-511 [FREE Full text] [doi: 10.1186/s12913-017-2523-6] [Medline: 28821243]
46. Zurovac D, Otieno G, Kigen S, Mbithi AM, Muturi A, Snow RW, et al. Ownership and use of mobile phones among health workers, caregivers of sick children and adult patients in Kenya: cross-sectional national survey. Global Health 2013 May 14;9:20 [FREE Full text] [doi: 10.1186/1744-8603-9-20] [Medline: 23672301]
47. Labrique A, Blynn E, Ahmed S, Gibson D, Pariyo G, Hyder AA. Health surveys using mobile phones in developing countries: automated active strata monitoring and other statistical considerations for improving precision and reducing biases. J Med Internet Res 2017 May 05;19(5):e121 [FREE Full text] [doi: 10.2196/jmir.7329] [Medline: 28476726]
48. Mahfoud Z, Ghandour L, Ghandour B, Mokdad AH, Sibai AM. Cell phone and face-to-face interview responses in population-based surveys. Field Methods 2014 Jul 01;27(1):39-54. [doi: 10.1177/1525822X14540084]
49. Lau C, di Tada N. Identifying non-working phone numbers in Africa for response rate calculations. Surv Pract 2017 Dec 28:1-3. [doi: 10.29115/sp-2018-0020]
50. Guwatudde D, Mutungi G, Wesonga R, Kajjura R, Kasule H, Muwonge J, et al. The epidemiology of hypertension in Uganda: findings from the National Non-Communicable Diseases Risk Factor Survey. PLoS One 2015 Sep 25;10(9):e0138991-e0138913 [FREE Full text] [doi: 10.1371/journal.pone.0138991] [Medline: 26406462]

Abbreviations

BDT: Bangladeshi Taka
RR: risk ratio
UGX: Ugandan Shilling

XSL•FO
RenderX

Edited by A Mavragani; submitted 31.01.22; peer-reviewed by K L'engle, IV George; comments to author 28.02.22; revised version received 10.03.22; accepted 31.03.22; published 09.05.22.

Please cite as:
Gibson DG, Kibria GMA, Pariyo GW, Ahmed S, Ali J, Labrique AB, Khan IA, Rutebemberwa E, Flora MS, Hyder AA
Promised and Lottery Airtime Incentives to Improve Interactive Voice Response Survey Participation Among Adults in Bangladesh and Uganda: Randomized Controlled Trial
J Med Internet Res 2022;24(5):e36943
URL: https://www.jmir.org/2022/5/e36943
doi:10.2196/36943
PMID:35532997

©Dustin Garrett Gibson, Gulam Muhammed Al Kibria, George William Pariyo, Saifuddin Ahmed, Joseph Ali, Alain Bernard Labrique, Iqbal Ansary Khan, Elizeus Rutebemberwa, Meerjady Sabrina Flora, Adnan Ali Hyder. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.05.2022. 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.

XSL•FO
RenderX

Original Paper

Self-administered Web-Based Tests of Executive Functioning and Perceptual Speed: Measurement Development Study With a Large Probability-Based Survey Panel

Ying Liu[1*], PhD; Stefan Schneider[1,2,3*], PhD; Bart Orriens[1], PhD; Erik Meijer[1], PhD; Jill E Darling[1], MSHS; Tania Gutsche[1], BA; Margaret Gatz[1,3], PhD

[1]Center for Economic and Social Research, University of Southern California, Los Angeles, CA, United States

[2]Leonard Davis School of Gerontology, University of Southern California, Los Angeles, CA, United States

[3]Department of Psychology, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Ying Liu, PhD
Center for Economic and Social Research
University of Southern California
635 Downey Way
Los Angeles, CA, 90089
United States
Phone: 1 2138211850
Fax: 1 2138212716
Email: liu.ying@usc.edu

Abstract

Background: Cognitive testing in large population surveys is frequently used to describe cognitive aging and determine the incidence rates, risk factors, and long-term trajectories of the development of cognitive impairment. As these surveys are increasingly administered on internet-based platforms, web-based and self-administered cognitive testing calls for close investigation.

Objective: Web-based, self-administered versions of 2 age-sensitive cognitive tests, the Stop and Go Switching Task for executive functioning and the Figure Identification test for perceptual speed, were developed and administered to adult participants in the Understanding America Study. We examined differences in cognitive test scores across internet device types and the extent to which the scores were associated with self-reported distractions in everyday environments in which the participants took the tests. In addition, national norms were provided for the US population.

Methods: Data were collected from a probability-based internet panel representative of the US adult population—the Understanding America Study. Participants with access to both a keyboard- and mouse-based device and a touch screen–based device were asked to complete the cognitive tests twice in a randomized order across device types, whereas participants with access to only 1 type of device were asked to complete the tests twice on the same device. At the end of each test, the participants answered questions about interruptions and potential distractions that occurred during the test.

Results: Of the 7410 (Stop and Go) and 7216 (Figure Identification) participants who completed the device ownership survey, 6129 (82.71% for Stop and Go) and 6717 (93.08% for Figure Identification) participants completed the first session and correctly responded to at least 70% of the trials. On average, the standardized differences across device types were small, with the absolute value of Cohen *d* ranging from 0.05 (for the switch score in Stop and Go and the Figure Identification score) to 0.13 (for the nonswitch score in Stop and Go). Poorer cognitive performance was moderately associated with older age (the absolute value of *r* ranged from 0.32 to 0.61), and this relationship was comparable across device types (the absolute value of Cohen *q* ranged from 0.01 to 0.17). Approximately 12.72% (779/6123 for Stop and Go) and 12.32% (828/6721 for Figure Identification) of participants were interrupted during the test. Interruptions predicted poorer cognitive performance ($P<.01$ for all scores). Specific distractions (eg, watching television and listening to music) were inconsistently related to cognitive performance. National norms, calculated as weighted average scores using sampling weights, suggested poorer cognitive performance as age increased.

XSL•FO
RenderX

Conclusions: Cognitive scores assessed by self-administered web-based tests were sensitive to age differences in cognitive performance and were comparable across the keyboard- and touch screen–based internet devices. Distraction in everyday environments, especially when interrupted during the test, may result in a nontrivial bias in cognitive testing.

(J Med Internet Res 2022;24(5):e34347) doi:10.2196/34347

KEYWORDS

cognitive tests; internet; probability-based; web-based; executive function; response speed; self-administered test; mobile phone

Introduction

Background

Internet-based surveys have received widespread attention as methods of large-scale data collection in many fields of health research. Although internet surveys have traditionally focused on the collection of self-reported data (eg, participants' subjective attitudes, health behaviors, and self-reported medical conditions), interest in the ability to conduct objective testing of cognitive abilities over the internet has substantially increased in recent years [1-4]. Cognitive capacities are relevant to people's ability to understand and act on information and serve as a basis for higher-order functioning and well-being. Large-scale monitoring of cognitive abilities is necessary to determine incidence rates, risk factors, and long-term trajectories of the development of cognitive impairment associated with chronic conditions [5] and normal cognitive aging [6].

Tests of cognitive functioning are often included in large population surveys administered face to face or over the telephone [7,8]. Compared with these conventional assessment strategies, large-scale internet-based cognitive testing has many potential benefits. Internet surveys have the obvious advantages of lower labor costs and quicker turnaround while achieving demographic representations of the population similar to those of traditional surveys [9]. Similar to conventional computerized testing where a psychometrist is present [10], web-based cognitive testing achieves higher precision and data quality because of electronic scoring compared with human-based scoring (eg, stopwatch), especially in timed assessments. Given that web-based cognitive tests are self-administered, do not require an examiner to be present, and eliminate manual entry of data, they can be much more flexibly and efficiently administered compared with conventional cognitive test formats [11]. Participants can be tested at various times on their own computers or mobile devices in their homes or in other daily environments; responses can be automatically routed and test responses stored electronically, thus making data available to researchers or other interested stakeholders in near real time. This allows for data collection with much larger and more diverse participant samples [3,12].

Notwithstanding the potential advantages of web-based cognitive testing, there are many open questions about the accuracy of its administration. Self-administered web-based cognitive tests reduce the level of standardization often seen in tests administered by trained professionals, which involves precisely controlled test environments and standardized equipment. The lack of standardization is particularly fraught for tests that are timed. Cognitive functioning scores may potentially differ markedly depending on participants' computer skills and the used device [13,14], such as whether the test is taken on a laptop computer with a physical keyboard versus a tablet or smartphone without a keyboard [1,15]. Moreover, although self-administered web-based testing increases flexibility and convenience of test administration in participants' daily lives, this comes with the potential costs of excessive measurement errors and biases associated with environmental influences (eg, the location where the test is taken and interruption by people or devices). To mitigate potential biases, it has been recommended that web-based cognitive tests should be specifically designed for unmonitored settings with clear instructions and consistent administration across operating systems, browsers, and devices [16]. Furthermore, as normative data may not be comparable across different modes of administration, norms should be made accessible specifically for web-based cognitive testing and generated from large and representative samples [13]. To date, available web-based cognitive tests have been predominantly developed for detecting age-related cognitive impairments in small and selected samples and lack normative information that could be applied in population-based studies [1,3].

To facilitate large-scale cognitive testing in the digital era, we developed web-based self-administered tests of perceptual speed and executive functioning in an existing US representative internet survey panel—the Understanding America Study (UAS). Panelists in the UAS complete monthly surveys and assessments on their own devices in everyday environments, providing an opportunity to evaluate the performance of these measures in a large and diverse sample. This allowed us to explore the extent to which the use of different web-based devices and environmental influences (eg, location or distractions) affects the precision of cognitive test scores and develop nationally representative norms for use in population-based research. Many web-based surveys and panels are opt-in samples recruited on the web, such as through Facebook advertisements. Using such a sample would potentially risk overrepresenting individuals with stronger connections to digital technology who are younger and have a higher socioeconomic status. To avoid such biases, the UAS recruits panel members using traditional methods (sampling addresses from the United States Postal Service Delivery Sequence File and initially contacting sampled individuals by postcards and letters) and provides them with a tablet and internet connection if they do not have access to the internet.

We focus on 2 cognitive domains—perceptual speed and executive functioning—that involve the speed of answering questions. Perceptual speed refers to how long an individual requires to take in information, understand it, and begin to act on it. It is typically measured as the time required to perceive

XSL•FO
RenderX

and respond to visual information. Executive functioning entails the cognitive skills used to control behavior. Key executive functions include response inhibition (resisting automatic impulses to act), interference control (suppressing attention to unwanted information), and mental set shifting (being able to change perspectives) [17]. Both perceptual speed and executive functioning are fundamental cognitive skills that are distinct from other cognitive domains such as verbal, spatial, and memory abilities in that they are more sensitive to stress, depression, lack of sleep, and poor physical health [17-19]. Thus, they can be important mediators in health-related studies [20]. Although performance can be improved with training and practice [17,21], these skills decline with age [22] and serve as critical indicators of healthy aging or potential flags of impairment. Declines in perceptual speed have also been shown to portend other cognitive changes in older adults [23]. Furthermore, their time-dependent nature and proneness to environmental influences make them sensitive to errors or biases potentially induced by platform differences or everyday environments.

Objectives of This Study

Our study had 3 objectives. The first was to examine differences in cognitive test scores across different web-based devices and the relationships between cognitive test scores and participant age. To do so, we asked the UAS panelists to complete the cognitive tests both on a keyboard-based computer (desktop or laptop) and on a device with a touch screen (tablet or smartphone) in a randomized order if both device types were available or complete the same test twice if only one type of device was available. The second objective was to study the extent to which test scores are related to environmental influences in everyday life, such as the location at which the test occurs and momentary distractions. The third objective was to provide test score norms based on a nationally representative sample in which individuals with cognitive impairment or potential dysfunction were not screened in or out. It is worth noting that there is no clinical gold standard; thus, these norms are not intended for clinical purposes, and thus, cognitive impairment cutoffs are outside the scope of our work.

Methods

Participants

The data for this study were collected as part of the UAS [24], a probability-based internet panel maintained at the University of Southern California [25]. In contrast to convenience (*opt-in*) panels, where participants self-select as members, UAS panelists are recruited through nationwide address-based sampling. This recruitment strategy relies on samples drawn with a known probability of selection from a US Postal Service list of all households in the nation, which tends to overcome many biases in population parameters estimated from convenience panels [26,27]. Participants without prior internet access are equipped with broadband internet and a tablet, which is important, given that internet access tends to be lower among older Americans and those with lower education [28]. UAS members are asked to complete 1 to 2 web-based assessments per month on various topics, including psychosocial well-being, economic concerns, retirement planning, decision-making strategies, and cognitive assessments. All active members of the UAS internet panel were asked to participate in these assessments, and no specific inclusion or exclusion criteria were applied for participation in this study.

The participants provided electronic informed consent for participation.

Measures

Executive Functioning

We selected the Stop and Go Switching Task (SGST) developed for telephone administration by Lachman et al [7,29], which was implemented in the Midlife in the United States national longitudinal study. In the original phone-administered version, the experimenter says the word *red* or *green,* and the participant responds by saying either *stop* or *go*. The SGST comprises several conditions that are administered in series. In the normal condition, the participant responds *stop* to *red* and *go* to *green*. It is followed by a reverse condition where the participant responds *go* to *red* and *stop* to *green*. These 2 *baseline* conditions are followed by a mixed condition in which participants switch back and forth between normal and reverse instructions. The switch trials are the first response after the participant has to change from one condition to another. Nonswitch trials are those that do not involve a change in instructions. The participants practice each condition before beginning. Then, latencies are measured (based on audio recordings of the telephone assessments) between the cue and the response for the normal, reverse, switch, and nonswitch trials. The median response time in each type of trial is used as a score for one's cognitive ability. The baseline normal condition measures choice reaction time, the reverse condition requires response inhibition, and the mixed condition requires task switching all of which are aspects of executive functioning [7].

Perceptual Speed

The test of perceptual speed that we selected was the Figure Identification test, which was originally developed as a paper-and-pencil test. This is based on the work of Thurstone [30] on primary mental abilities. The participant sees a target figure on top of 5 horizontally aligned similar figures. All figures are in black and white and vary in complexity, with some but not all representing recognizable objects (eg, an abstract dog or boat). The task is to identify 1 figure among the 5 that exactly matches the target as quickly as possible while being accurate. Perceptual speed is measured by counting the number of figures correctly circled on paper within a preset time limit. The paper-and-pencil version of the Figure Identification test has long been used as part of the Dureman and Sälde battery [31], especially in studies of cognitive aging [32].

Web-Based Adaptation of the Cognitive Tests

Prior Adaptation for the Web

Developers had already taken steps to develop web-based versions of both cognitive tests. As much as possible, they emulated the original tests, except that the stimuli were presented on an electronic visual display (rather than on paper or via telephone) and responded to by pressing keys on a

XSL•FO
RenderX

keyboard or buttons on a touch screen (rather than circling responses on paper or responding verbally), and responses and reaction times were electronically captured. For the Figure Identification test, respondents pressed the correct answer rather than circling the correct answer on a sheet of paper showing the figures (personal communication, Johansson). For SGST, participants view the word *red* or *green* and respond by pressing the *S* (stop) or *G* (go) key rather than answering verbally. Viewing words rather than viewing, for example, a red or green disk, eliminates the issues of color blindness. To minimize the motor component in response time, participants were encouraged to keep their fingers on the keys (JJ McArdle, CA Prescott, EE Walters, GG Fisher, B Helppie McFall, K Peters, unpublished user documentation, May 15, 2018). Administration of the version, as developed by JJ McArdle, CA Prescott, EE Walters, GG Fisher, B Helppie McFall, and K Peters (unpublished user documentation, 2018), to a sample of 408 participants who completed the SGST both by phone and web found longer response times for web than for phone for normal and reverse baseline conditions, longer response times for phone than for web for switch trials, and no difference for nonswitch trials (R McCammon, personal communication, January 11, 2022).

Further Adaptation in This Study

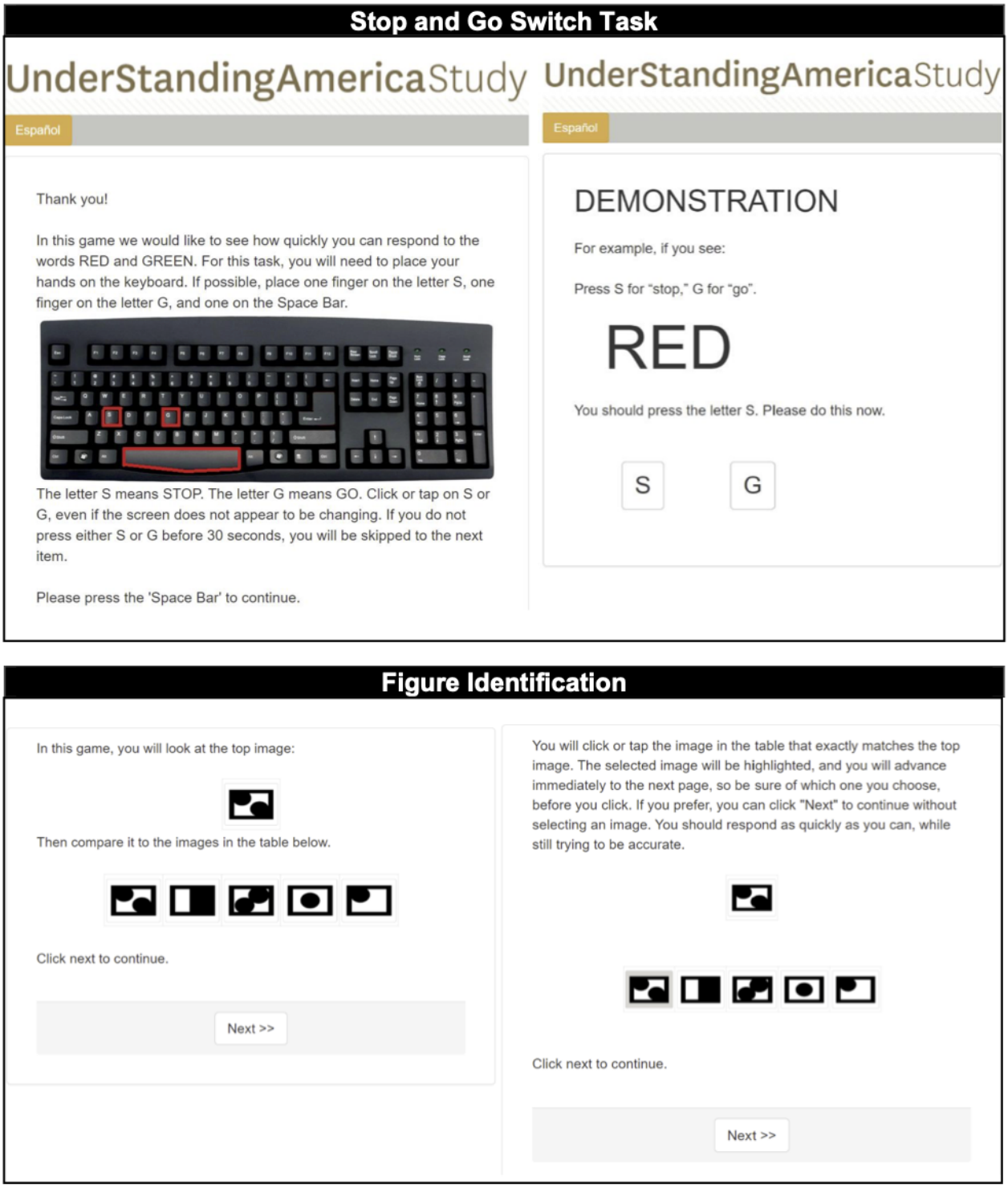
The UAS team administers surveys using the NubiS data collection tool, an open-source, secure data collection, storage, and dissemination system [33] developed at the Center for Economic and Social Research, University of Southern California. Surveys in NubiS are conducted in a web browser environment designed to optimize the harmonization of the survey experience across a wide variety of devices and browsers. This avoids the need to accommodate changes in the device or the web browser environment. Given the specific user interaction mechanisms of the web-based cognitive tests, the surveys were further refined to be administered on both devices with a keyboard and mouse (eg, desktop or laptop computers) and touch screen-based devices (eg, tablets or smartphones). For keyboard-based devices, the interface responds to keys pressed on the keyboard or mouse clicks. For touch screen–based devices, the interface incorporates buttons for the possible answer keys, which, when pressed, simulate the behavior of their respective keyboard or mouse counterparts.

Figure 1 presents selected screenshots of the web-based versions. At the beginning of each test, participants were given a brief introduction to the task (see the left panels in Figure 1), followed by a demonstration (right panels). After the demonstration, practice trials were provided with automated feedback before participants were asked to start the test. In the feedback, participants were told whether their answer was correct or incorrect; if incorrect, they were shown the correct answer and were given another practice trial. The participants were also instructed to set aside the uninterrupted time to complete the tasks.

For both tests, item latencies were recorded in the client browser as time lapsed, in milliseconds, between the moment of *screen fully loaded* to the moment that an answer key was pressed or the button was clicked (captured through a JavaScript *onkeyup* event). The event of a fully loaded screen or page is captured by a JavaScript document-ready expression from the client's browser. In this way, browser speed differences displaying the page are excluded from the latency data, as is the time spent on the server-client interaction. This ensures that any differences in platform or browser speed in displaying or internet speed do not affect the recorded response latencies.

XSL•FO
RenderX

Figure 1. Screenshots of the introductory (left) and demonstration (right) pages of web-based versions of the Stop and Go Switching Task (top) and the Figure Identification test (bottom).



Pilot Testing of the Web-Based Versions

A combined total of 964 UAS panelists were initially recruited for a sequence of pilot tests and to provide feedback about their experiences while performing the web-based tests. Feedback was provided by respondents at the end of the pilot surveys via closed-ended questions about the usability of various devices, clarity of instructions, and an open-ended text box for additional comments. The authors (JD, SS, YL, EM, and MG) used descriptive statistics to examine responses to the feedback questions and indicators such as item and survey nonresponse, break off, and level of compliance with written directions. Several refinements were made during the iterative process, informed by the pilot data and participant feedback. First, the number of practice sessions for the SGST was increased from 1 to 3 practice sessions for the baseline and reverse conditions after 2 initial demonstrations to ensure that all participants were fully aware of the specifics of the test for all trials. Second, the wording of the instructions was refined to increase clarity in

XSL•FO
RenderX

response to participant feedback. Third, we simplified the presentation and layout on the screen by eliminating superfluous elements (eg, a *refuse* button) on the screen. We also minimized the number of different keys and streamlined the number of hand motions that the participants needed to use throughout the tests. For example, in the SGST, when a respondent is asked to keep their fingers on certain keys on the keyboard, using the spacebar to move to the next page is more natural than pressing the *next* button, which requires operating a mouse (see the top left panel in Figure 1).

Main Study Procedures and Measures

Overview

To compare the test performance when using a device with a keyboard and mouse versus a touch screen, we designed a study in which participants were asked to complete the tests twice in separate modules. Those who had access to both device types were randomized into using a keyboard and mouse first and touch screen second, or vice versa, whereas participants who had access to only 1 of the 2 types were asked to complete the tests twice on the same device. Leveraging the within-person design resulted in a much reduced and nonrepresentative sample (1770 instead of 6129 for SGST and 1892 instead of 6717 for the Figure Identification test). Therefore, our primary analyses were based on the participants' first completed test (ie, based on a between-person comparison across device types). However, we also report results from sensitivity analyses conducted in a reduced sample of participants who provided data for the within-person comparison across device types, which is in line with the original study design.

The types of devices used by participants were monitored using Mobile Detect, a hypertext preprocessor–based tool for analyzing and classifying browser user agent strings [34]. When entering each test session, the tool automatically analyzed the browser user agent string to determine the device type, and the survey instructed the participants to confirm or switch devices if the detected type was inconsistent with the assigned type. The SGST and Figure Identification tests were administered separately, which were an average of 10.6 (SD 53.1) days apart.

SGST Scoring

The SGST included 10 normal and 10 reverse baseline trials, as well as 23 nonswitch and 6 switch trials. Following prior research using the telephone-administered version [7], each of these 4 types of tasks was scored using median response times across all trials within each trial type to reduce the effects of outlier responses. Participants who failed to meet an acceptable level of overall accuracy (at least 70% correct trials) were not scored. Higher scores indicate slower median response times and, thus, poorer executive functioning. The test items were simple enough that anyone could answer them correctly when spending sufficient effort. The response times then measured the concepts of interest as the amount of effort required to respond. When a respondent answers many items incorrectly, it, therefore, likely indicates problems such as carelessness or inattention, which may correspond to rushing through the items, making the response times less valid measures of the concepts we intended to measure. To guard against this, we followed the literature and excluded respondents with too many incorrect items. There is a trade-off between validity at the individual level and representativeness and the overall sample size. We considered several cutoffs and, based on descriptive analyses of early data extracts, concluded that 70% best balanced the 2 objectives in this trade-off.

Figure Identification Test Scoring

In the original paper-and-pencil version, the Figure Identification test was administered in 2 sets of 30 items each. Participants were given 120 seconds per set to complete as many items as they could, and the final score was the total number of figures that a person had solved correctly within the preset time limits [31]. In this study, administering the test on the web made it unnecessary to terminate the test after a preset time limit, as we were able to set posterior time limits [35]. Therefore, we asked participants to complete all 60 figures. To eliminate potential order effects, the figures were grouped into 6 blocks of 10, and the order of the blocks was randomized. We used the median and IQR of the item-level response times obtained from the pilot data as proxies of item difficulty and discrimination measures when creating the blocks of items. The 6 items with the lowest median were randomly assigned to the 6 blocks, and so were the remaining items. To ensure the resulted blocks were comparable, we repeated this process 500 times and chose a combination with a similar overall median and IQR across blocks.

To derive a final Figure Identification test score, we used a posterior time limit of 90 seconds per 30-item set as preliminary analyses indicated ceiling effects when the original 120-second limit was used, with many participants obtaining scores approaching the maximum possible score. It is well-known that electronic responses tend to be faster than paper-based because of differences in item presentation and response format (eg, pressing a key instead of marking a mark with a pen) [36]. Thus, the final score was the total number of correctly identified figures within 2×90 seconds. As with the SGST, we required that a respondent had at least 70% correct figures to be scored. The rationale was to screen out inattentive responders as our prior research has shown that having <70% figures incorrect on the Figure Identification test is an indicator of careless, inattentive responding and likely yields invalid test results [37]. Extreme outlier response times >30 seconds (99.80th percentile across all responses) were removed from the data and did not count toward the respondent's correctly solved figures. Higher scores on the test reflect a faster perceptual speed.

Although, in combination, the scoring of the Figure Identification test reflects both the speed and accuracy of responses, it is also of interest to examine these components separately. To date, it is not known whether device differences or environmental influences in self-administered web-based cognitive tests affect response speed, response accuracy, neither, or both. Therefore, as a secondary set of scores derived from the test, we also examined (1) response inaccuracy (ie, the percentage of figures a respondent had incorrect) and (2) participants' response latency (ie, the median response time of all accurate trials) as separate outcome measures. For both measures, higher scores reflected worse performance.

XSL•FO
RenderX

Posttest Survey on Readability and Screen Navigation

After each test session, the participants were asked about their experiences. This included (1) how clearly they were able to see the text, buttons, or figures (very clearly, not very clearly, and not clearly at all); (2) difficulty navigating the screen such as tapping, clicking, or needing to scroll (very easy, somewhat easy, somewhat difficult, and very difficult); and (3) the overall experience (eg, needing to scroll or move the screen to see all content, content obscured by something else, or no technical difficulties). They were also provided an open-ended text box to leave additional comments.

Environmental Influences

Information on participants' current location and momentary distractions was collected via self-report immediately after each session of the cognitive test. A single item asked the participants where they were when they completed the test (at home or their residence, at work, at school, in a public place, riding a car or other transportation, or walking outside), which was recoded as being at home (1) or not (0). To assess momentary distractions, participants were asked about other activities they performed while they completed the test (talking to other people, listening to music or podcasts, watching television, playing games, following content on the internet, texting, or checking their email). Each activity option was coded as yes (1) or no (0). Finally, participants were asked whether they were interrupted by anything while completing the test (answers: yes or no).

Demographic Variables

Demographic information was collected quarterly from the UAS survey panel. This included sex, age, race and ethnicity, education, and household income. The mean age of participants in this study was 49.6 (SD 15.8, range 18-101) years for the analysis of the SGST and 50.5 (SD 16.0, range 18-102) years for the Figure Identification test. Approximately three-fifths were female (3620/6129, 59.06% for SGST; 4021/6717, 59.86% for Figure Identification), approximately two-thirds were White (4030/6119, 65.86% for SGST; 4323/6708, 64.45% for Figure Identification), slightly less than half had a bachelor's degree or more (2732/6127, 44.59% for SGST; 2797/6715, 41.65% for Figure Identification), and approximately three-fifths had a household income of ≥US $50,000 (3730/6113, 61.01% for SGST; 3922/6701, 58.53% for Figure Identification).

Analysis Strategy

The analyses were conducted separately for the collected SGST and Figure Identification test data. For primary analyses of each cognitive test, we included all participants who completed at least one session of the test and responded correctly to at least 70% of the trials, regardless of the type of device used. Sensitivity analyses examined within-person differences in test scores among the smaller group that completed the tests on both device types. Demographic characteristics were compared across device ownership groups using the Pearson chi-square test. A comparison of the cognitive scores across the keyboard- and touch screen–based devices was performed in 2 ways. First, we used Cohen *d* as a measure of the standardized mean difference. We considered the standardized mean difference to be small with d=0.20, medium with d=0.50, and large with d=0.80 [38]. Second, Pearson correlation was used to quantify the association between the cognitive scores and the respondent's age, where scatterplots were inspected to detect potential nonlinear relationships. Cohen *q* was used to quantify differences in the age correlation across device types, where values of 0.10, 0.30, and 0.50 can be interpreted as small, medium, and large effects, respectively [38]. To understand the association between cognitive scores and environmental influences, multivariable regression was used with age as a covariate. We controlled for age in this analysis to avoid omitted variable bias, given that cognitive performance is hypothesized to decline as one's age increases, and environmental influences may also be related to age. Effects at P<.05 were considered statistically significant. Finally, US population norms for cognitive scores were computed as weighted mean cognitive scores for different age groups using standard UAS survey weights [39]. All analyses were performed using Stata (version 16) and SAS (version 9.4).

Ethics Approval

This study was approved by the institutional review board of the University of Southern California (UPS 14-00148).

Results

Descriptive Characteristics of the Analysis Sample

For the SGST, among the 9453 panelists invited to participate, 7410 (78.39%) completed the device ownership survey. Of those 7410 individuals, 7039 (94.99%) completed the first session, and 6129 (82.71%) correctly responded to at least 70% of the trials and received a score; one case was excluded from the analysis because of a lack of information about the device they used. For the Figure Identification test, 9445 were invited, and 7216 (76.4%) participants completed the device ownership survey. Of those 7216 individuals, 6879 (95.33%) completed the first session, and 6717 (93.08%) correctly identified at least 70% of the figures and were analyzed. A small proportion of the participants (194/7292, 2.66% for SGST; 9/6888, 0.13% for the Figure Identification test) dropped out after the practice trials.

Table 1 shows the demographic composition of the analytic samples overall and by the devices owned. Those with only a touch screen device were younger, had lower education, had lower household income, and were less likely to be male or White than those with only a keyboard device. Individuals who had both types of devices tended to have higher education and household income and were slightly younger and more likely to be White than single-device owners.

XSL•FO
RenderX

Table 1. Demographic characteristics of the study samples.

Participant characteristics	Stop and Go Switching Task (n=6129)			*P* value	Figure Identification test (n=6717)			*P* value
	Both devices, n (%) (n=3228)	Keyboard only, n (%) (n=959)	Mobile only, n (%) (n=1942)		Both devices, n (%) (n=3345)	Keyboard only, n (%) (n=1172)	Mobile only, n (%) (n=2200)	
Age (years)				<.001				<.001
18-34	687 (21.28)	99 (10.3)	443 (22.81)		684 (20.45)	100 (8.53)	475 (21.59)	
35-44	688 (21.31)	129 (13.5)	474 (24.41)		703 (21.02)	156 (13.31)	520 (23.64)	
45-54	623 (19.3)	154 (16.1)	365 (18.8)		635 (18.98)	165 (14.08)	433 (19.68)	
55-64	591 (18.31)	226 (23.6)	368 (18.95)		643 (19.22)	280 (23.89)	408 (18.55)	
65-74	501 (15.52)	239 (25)	225 (11.59)		516 (15.43)	293 (25)	287 (13.05)	
≥75	134 (4.15)	110 (11.5)	67 (3.45)		164 (4.9)	178 (15.19)	77 (3.5)	
Race				<.001				<.001
Non-Hispanic White	2155 (66.76)	692 (72.3)	1183 (60.92)		2184 (65.29)	850 (72.53)	1289 (58.59)	
Non-Hispanic Black	210 (6.51)	46 (4.8)	167 (8.6)		237 (7.09)	72 (6.14)	204 (9.27)	
Hispanic	490 (15.18)	115 (12)	422 (21.73)		519 (15.52)	130 (11.09)	514 (23.36)	
Non-Hispanic other	367 (11.37)	104 (10.9)	168 (8.65)		400 (11.96)	119 (10.15)	190 (8.64)	
Sex				<.001				<.001
Men	1332 (41.26)	482 (50.3)	695 (35.79)		1342 (40.12)	604 (51.54)	750 (34.09)	
Women	1896 (58.73)	477 (49.7)	1247 (64.21)		2003 (59.88)	568 (48.46)	1450 (65.91)	
Education				<.001				<.001
High school or less	433 (13.41)	203 (21.2)	592 (30.48)		478 (14.29)	267 (22.78)	724 (32.91)	
Some college	1028 (31.85)	322 (33.6)	817 (42.07)		1113 (33.27)	415 (35.41)	921 (41.86)	
Bachelor or more	1766 (54.71)	434 (45.3)	532 (27.39)		1754 (52.44)	490 (41.81)	553 (25.14)	
Household income (US $)				<.001				<.001
≤24,999	396 (12.27)	159 (16.6)	541 (27.86)		453 (13.54)	216 (18.43)	671 (30.5)	
25,000-49,999	573 (17.75)	233 (24.4)	481 (24.77)		598 (17.88)	283 (24.15)	558 (25.36)	
50,000-99,999	1139 (35.29)	343 (35.9)	537 (27.65)		1167 (34.89)	420 (35.84)	579 (26.32)	
≥100,000	1109 (34.36)	221 (23.1)	381 (19.62)		1119 (33.45)	247 (21.08)	390 (17.73)	

Readability and Test Experience

Of the 6129 and 6717 participants who met the 70% correctness threshold to be in the analysis samples for the SGST and the Figure Identification tests, 5901 of 6129 (96.28 for SGST) and 6492 of 6715 (96.68% for the Figure Identification test) reported seeing the text and buttons very clearly, 5746 of 6129 (93.75% for SGST) and 6372 of 6714 (94.91% for the Figure Identification test) considered it very or somewhat easy to navigate the screen, and 5231 of 6129 (85.35% for SGST) and 6075 of 6710 (90.54% for the Figure Identification test) reported that the text and buttons fit on the same screen with no technical difficulties, respectively. Those who completed the session but did not meet the 70% correctness threshold reported somewhat more problems. For the SGST, 73.7% (671/910) reported seeing the content very clearly, 72.7% (662/910) considered it very or somewhat easy to navigate, and 66% (598/906) had no technical difficulties; for the Figure Identification test, 63.8% (88/138) reported seeing the content very clearly, 70.8% (97/137) considered it very or somewhat easy to navigate, and 54.8% (74/135) had no technical difficulties.

Comparison of Cognitive Scores Across Device Types

The results of the primary analyses of SGST are shown in the top panel of Table 2. The average SGST scores, which are the median response times for each type of trial, ranged from 0.92 to 1.53 seconds. The switch trials were the slowest, and the nonswitch trials were the fastest. The difference across device types was small on average, with scores on touch screen devices being 0.03 to 0.06 seconds slower than on keyboard devices. Standardized mean differences (Cohen *d*) ranged from 0.05 (for switch trials) to 0.13 (nonswitch trials), which is below the threshold of 0.20 for a small effect. All 4 scores were positively correlated with age (correlations ranged from 0.32 to 0.50), indicating a slower performance with higher age. The age correlations were similar across platforms; Cohen *q* for differences in age correlations ranged in absolute value from 0.01 (for switch trials) to 0.09 (for reverse baseline trials), which is below the threshold of 0.10 for a small effect.

XSL•FO
RenderX

Table 2. Comparison of cognitive scores and their correlation with age across device types.

Cognitive scores	Values, mean (SD)		Cohen *d* (95% CI)	Correlation with age		Cohen *q*
	Keyboard	Touch screen		Keyboard	Touch screen	
Stop and Go Switching Task[a]						
Baseline[b]	1.04 (0.56)	1.09 (0.80)	0.07 (0.02 to 0.12)	0.38	0.32	0.07
Reverse baseline[b]	1.10 (0.58)	1.16 (0.59)	0.10 (0.05 to 0.15)	0.35	0.43	−0.09
Nonswitch[b]	0.92 (0.32)	0.97 (0.34)	0.13 (0.08 to 0.18)	0.48	0.50	−0.03
Switch[b]	1.50 (0.70)	1.53 (0.63)	0.05 (−0.01 to 0.10)	0.35	0.34	0.01
The Figure Identification test[c]						
The Figure Identification test score[d]	41.51 (8.43)	41.96 (8.36)	0.05 (0.01 to 0.10)	−0.61	−0.49	−0.17
Percentage figures incorrect[e]	5.70 (4.97)	7.40 (5.70)	0.32 (0.27 to 0.37)	−0.04	−0.05	0.00
Median response times[b]	4.76 (1.78)	4.53 (1.74)	−0.13 (−0.18 to −0.08)	0.56	0.45	0.15

[a]For Stop and Go Switching Task, keyboard n=2820 and touch screen n=3309.

[b]Means and SDs are presented for seconds.

[c]For the Figure Identification test, keyboard n=3182 and touch screen n=3309.

[d]Means and SDs are presented for the number of figures.

[e]Means and SDs are presented for percentage.

As shown in the bottom panel of Table 2, the average Figure Identification test scores were very similar across device types, with a mean score of 41.5 (SD 8.43, possible range 0-60) for keyboard and 41.96 (SD 8.36) for touch screen devices (d=0.05). The scores showed pronounced negative correlations with age (r=−0.61 when completed on a keyboard device and r=−0.49 when completed on a touch screen device). The age relationship was stronger for the keyboard than for the touch screen (q=−0.17, just less than halfway between a small and medium effect size). The results for the secondary Figure Identification outcome measures suggested that participants made more mistakes when using a touch screen device (average percent incorrect figures 7.4, SD 5.70) than when using a keyboard device (average percent incorrect figures 5.7, SD 4.97; d=0.32, just less than half way between a small and medium effect), whereas response times were faster on average when using a touch screen device (mean of median response times 4.53, SD 1.74) than when using a keyboard device (mean of median response times 4.76, SD 1.78; d=0.13, a small effect). Older age was weakly associated with a lower percentage of incorrect figures (r=−0.044 for keyboard and r=−0.045 for touch screen; q=0.001) and strongly associated with slower responses (r=0.56 for keyboard and r=0.45 for touch screen; q=0.15). No meaningful nonlinear trend was observed between age and SGST or the Figure Identification test scores.

As a sensitivity check, we also conducted a *within-person* comparison of test scores among the smaller subgroup of participants who completed the cognitive tests consecutively on both device types. This analysis included 46.44% (1770/3811) of SGST participants and 55.76% (1892/3393) of Figure Identification test participants who owned both device types, took the test twice using a keyboard and touch screen device, and met the 70% correctness threshold for both sessions. Multimedia Appendix 1 Table S1 shows that respondents who met these criteria were younger on average; had higher education; had a higher household income; were more likely White; and (for the SGST sample) more likely male compared with respondents who owned both device types but did not meet the criteria. As shown in Multimedia Appendix 2 Table S2, respondents included in the within-person comparison showed somewhat faster response times on the SGST and better scores on the Figure Identification test (for both device types) compared with the full analysis sample. Within-person analyses of mean differences between devices largely replicated the results in the full sample, with effect sizes that were somewhat more pronounced for SGST and slightly smaller for the Figure Identification test. Correlations of the cognitive scores across device types were moderate to large, ranging from 0.37 (baseline trials in SGST) to 0.83 (median response times in the Figure Identification test).

Environmental Influences

Participants were allowed to take the cognitive tests at the time and location of their preferences. Hence, although they were instructed to set aside some uninterrupted time before the tests, it is possible that the participants experienced distractions during testing. Using self-reports on environmental factors, Table 3 shows that approximately 89.6% (5486/6123 for the SGST sample) and 89.57% (6020/6721 for the Figure Identification test) of the participants took the tests at home. Approximately 12% to 13% of participants (779/6122, 12.72% for the SGST sample; 828/6718, 12.33% for the Figure Identification test) reported being interrupted while completing the test. Watching television (930/6101, 15.24% for the SGST sample; 962/6696, 14.37% for the Figure Identification test), listening to music or podcasts (518/6101, 8.49% for the SGST sample; 576/6696, 8.6% for the Figure Identification test), and talking with others (398/6101, 6.52% for the SGST sample; 483/6696, 7.21% for

XSL•FO
RenderX

the Figure Identification test) were the most frequently reported distractors.

Tables 4 and 5 report the results of the regression analysis relating cognitive scores to potential distractions while controlling for age. Being interrupted during the SGST task was associated with significantly (P<.01) poorer performance on all subtests, with median response times being 0.07 (nonswitch trials) to 0.15 (switch trials) slower for participants who reported being interrupted. Watching television was predictive of slower responses on the nonswitch (0.03 seconds slower; P=.02) and switch trials (0.06 seconds; P=.007), and texting or checking email was predictive of slower responses on switch trials (0.23 seconds; P=.02). For the Figure Identification test, being interrupted during the task was associated with lower scores (0.89 fewer figures identified during the time limit; P=.002), as were watching television (1.44 fewer figures identified; P<.001), texting or checking email (2.87 fewer figures; P<.001), and playing another game during the task (4.21 fewer figures; P=.047). The results for the secondary Figure Identification test outcomes showed that watching television and following content on the internet were associated with a higher percentage of incorrect figures, whereas being interrupted, watching television, and texting or checking email were associated with slower median response times.

We also examined the association between environmental influences and age. If environments and distractions varied by age, this could at least partially account for the observed age differences in cognitive test scores. Younger age was significantly associated with a greater likelihood of taking the test away from home (r=.12, P<.001, both for the SGST and the Figure Identification test samples), being interrupted (r=.04, P<.001, SGST; r=.05, P<.001, Figure Identification test), talking (r=.09, P<.001, SGST; r=.12, P<.001, Figure Identification test), listening to music (r=.08, P<.001, SGST; r=.08, P<.001, Figure Identification test), following content on the internet (r=.05, P<.001, SGST; r=.04, P=.001, Figure Identification test), and texting or checking email during the test (r=.03, P=.008, Figure Identification test only). However, these factors did not meaningfully affect the relationship between age and cognitive scores. When comparing zero-order correlations between age and cognitive test scores with partial correlations that controlled for environmental influences, the correlations differed by less than q=0.01 for all tests.

Table 3. Frequency distribution of environmental influence factors.

Environmental influences	SGST[a] (n=6123), n (%)	Figure Identification test (n=6721), n (%)
Interrupted during test	779 (12.72)	828 (12.32)
Being at home[b]	5486 (89.6)	6020 (89.57)
Watching television	930 (15.19)	962 (14.31)
Listening to music or podcast	518 (8.46)	576 (8.57)
Talking to others	398 (6.69)	483 (7.19)
Texting or checking email	42 (0.69)	45 (0.67)
Following content on internet	41 (0.67)	41 (0.61)
Playing another game	7 (0.11)	11 (0.16)

[a]SGST: Stop and Go Switching Task.

[b]Alternative responses to being at home included taking the test at work, at school, in a public place, riding a car or other transportation, or walking outside.

Table 4. Regression results associating environmental distractors with the Stop and Go Switching Task performance (N=6095).

Environmental influences	Normal baseline		Reverse baseline		Nonswitch		Switch	
	b (SE)	*P* value	*b* (SE)	*P* value	*b* (SE)	*P* value	*b* (SE)	*P* value
Interrupted during test	0.08 (0.03)	.006	0.08 (0.02)	<.001	0.07 (0.01)	<.001	0.15 (0.03)	<.001
Being at home	0.01 (0.03)	.81	0.01 (0.02)	.54	0.02 (0.01)	.13	0.01 (0.03)	.61
Watching television	0.02 (0.02)	.50	0.03 (0.02)	.11	0.03 (0.01)	.02	0.06 (0.02)	.007
Listening to music or podcasts	0.03 (0.03)	.33	0.01 (0.03)	.77	0.00 (0.01)	.84	−0.01 (0.03)	.80
Talking to others	−0.03 (0.04)	.49	0.03 (0.03)	.35	0.00 (0.02)	.99	0.03 (0.04)	.44
Texting or checking email	0.09 (0.11)	.42	0.13 (0.09)	.13	0.00 (0.05)	.93	0.23 (0.10)	.02
Following content on the internet	0.02 (0.11)	.86	0.08 (0.09)	.34	−0.01 (0.05)	.83	0.00 (0.10)	.97
Playing another game	0.26 (0.25)	.29	−0.01 (0.21)	.97	0.11 (0.11)	.34	0.41 (0.24)	.08

XSL•FO
RenderX

Table 5. Regression results associating environmental distractors with the Figure Identification test performance (N=6687).

Environmental influences	Figure Identification test score		Percentage incorrect		Median response time	
	b (SE)	*P* value	*b* (SE)	*P* value	*b* (SE)	*P* value
Interrupted during test	−0.89 (0.28)	.002	.0.31 (0.22)	.16	0.18 (0.06)	.003
Being at home	−0.18 (0.29)	.54	−0.14 (0.22)	.53	0.07 (0.06)	.24
Watching television	−1.44 (0.25)	<.001	0.49 (0.19)	.01	0.23 (0.05)	<.001
Listening to music or podcasts	0.11 (0.31)	.72	−0.40 (0.24)	.09	0.08 (0.07)	.22
Talking to others	−0.60 (0.36)	.09	0.06 (0.28)	.82	0.09 (0.08)	.24
Texting or checking email	−2.87 (1.06)	.007	−1.70 (0.88)	.05	0.98 (0.23)	<.001
Following content on the internet	−0.19 (1.13)	.87	2.29 (0.88)	.01	0.05 (0.25)	.85
Playing another game	−4.21 (2.13)	.047	0.86 (1.65)	.60	0.60 (0.46)	.20

Norms

After applying the sampling weights developed for the UAS panel, we computed the norms for the two cognitive tests that were representative of the general US population. Tables 6 and 7 show the weighted averages of test scores by age group with 95% CIs, including participants who reported being distracted.

Table 6. Weighted averages of the Stop and Go Switching Task scores with 95% CI by age group (N=5933).

Age group (years)	Baseline average (95% CI)	Reverse baseline average (95% CI)	Nonswitch average (95% CI)	Switch average (95% CI)
18-34	0.80 (0.78-0.82)	0.88 (0.86-0.90)	0.76 (0.75-0.77)	1.25 (1.22-1.29)
35-44	0.90 (0.86-0.93)	0.98 (0.95-1.00)	0.84 (0.82-0.85)	1.36 (1.32-1.40)
45-54	1.04 (1.00-1.07)	1.11 (1.08-1.14)	0.94 (0.92-0.96)	1.55 (1.48-1.62)
55-64	1.29 (1.15-1.42)	1.28 (1.23-1.33)	1.04 (1.02-1.07)	1.61 (1.56-1.67)
65-74	1.44 (1.36-1.52)	1.50 (1.43-1.57)	1.20 (1.16-1.24)	1.86 (1.80-1.92)
≥75	1.50 (1.35-1.64)	1.49 (1.38-1.61)	1.22 (1.16-1.29)	2.01 (1.88-2.14)

Table 7. Weighted averages of the Figure Identification test scores with 95% CI by age group (N=6492).

Age group (years)	The Figure Identification test score (95% CI)	Percentage incorrect (95% CI)	Median response time (95% CI)
18-34	47.79 (47.27-48.31)	7.01 (6.64-7.39)	3.56 (3.48-3.64)
35-44	45.01 (44.45-45.56)	6.72 (6.34-7.11)	4.03 (3.93-4.14)
45-54	40.69 (40.10-41.28)	6.80 (6.34-7.25)	4.74 (4.62-4.88)
55-64	38.44 (37.89-38.98)	6.62 (6.19-7.04)	5.20 (5.06-5.34)
65-74	35.52 (35.04-36.00)	6.40 (5.95-6.86)	5.88 (5.73-6.03)
≥75	33.58 (32.80-34.36)	6.53 (5.83-7.23)	6.44 (6.18-6.71)

Discussion

Principal Findings

Surveys are increasingly being administered over the internet, posing questions about the quality of web-based information. This is especially true for measures of cognition. Cognitive tests have traditionally been administered in controlled environments under the supervision of a trained psychometrist, whereas administration in web surveys is potentially subject to spurious differences related to the type of device used by the respondent and distractions outside the control of the survey agency [13,16,40]. Nevertheless, assessing participants' cognitive abilities in large, nationally representative samples is often desirable [7,8]. In this paper, we studied web-based versions of 2 types of speeded cognitive tests—a switching test (SGST) to measure executive functioning and a matching test (Figure Identification test) to measure perceptual speed—in a nationally representative sample of US adults.

We developed the tests and their implementation iteratively through pilot tests and feedback from the participants in those pilots. Importantly, we imposed no restrictions on the system or hardware requirements, with the goal of broadly accommodating all devices that participants might have available. The final versions worked well for a large majority of participants in the full sample, and most participants reported experiencing no difficulties seeing the text and buttons clearly or navigating the screens. Somewhat greater difficulties with the self-administered tests were reported among the smaller subsets of participants who either did not finish the tests or provided <70% accurate answers and were, therefore, not scored.

XSL•FO
RenderX

It is also noteworthy that the rate of participants who did not meet this accuracy criterion was about twice as high (910/7039, 12.93%) for the SGST administered in this study compared with a previous report of telephone-administered SGST (262/4268, 6.1%) [7]. Further investigation revealed that 36.4% (332/912) of the excluded SGST sample failed in almost all the reverse baseline trials. This group's accuracy rate was high for the normal baseline trials as well as the practice trials of the reverse baseline condition. They also performed reasonably well in trials alternating between normal and reverse conditions (ie, switch and nonswitch trials), which suggests that these respondents might have mistakenly applied the normal baseline rules to the reverse baseline trials. For future respondents, we further modified the instructions by reiterating the reverse baseline rules between the practice and scorable trials of the same condition. As the development of web-based cognitive tests in the UAS is an ongoing process, further reduction of the remaining technical difficulties could continue to optimize test administration to ensure that the tests work as intended for all participants.

To compare any device effects, we asked individuals who had both a keyboard-based device and a touch screen–based device to perform the tests once on each device, for which we randomized the order. Successfully implementing this experimental study design component ultimately proved challenging as participants did not own both devices, did not agree to complete the tests on both devices, or did not use both devices as instructed, which resulted in a much reduced and nonrepresentative sample. This highlights the challenges frequently associated with executing randomized experiments in the context of large-scale internet panels [41]. Nevertheless, when comparing participants' scores for the first session, which yielded a very high participation rate, we found that keyboard- and touch screen-based devices yielded very similar scores in terms of participants' average cognitive performance. This was corroborated by the results from within-person analyses in the subsample of participants who successfully completed the experimental study design.

We found that older age was associated with worse scores on both cognitive tests, regardless of the device type. The observed worsening of scores was evident over the full adult age range, consistent with the theoretically expected age-normative cognitive trajectories [42]. Although we cannot rule out that the relationship with age is partially because of differences in familiarity with digital devices, the correlations between age and cognitive scores were consistent in magnitude with those previously reported for the original tests administered with traditional assessment formats. Scores on the SGST have been reported to correlate with age at 0.34 when the test was administered via telephone in the Midlife in the United States study [7], consistent with the correlations (ranging from 0.32 to 0.50) observed in this study. Similarly, age correlations for the traditional paper-and-pencil administered Figure Identification test ranged from −0.46 to −0.55 across waves in the Swedish Adoption/Twin Study of Aging [23,43], comparable in magnitude to those in this study (range −0.49 to −0.61). Furthermore, the tests themselves require only minimal familiarity with digital devices (pressing a specific key or button), which UAS members likely had already acquired in previous surveys in which they participated.

To date, only a few studies have examined the influence of different test settings on cognitive test scores [12,44]. To study the potential effects of the lack of a controlled environment, we asked the participants about their location and the number of potential distractions in day-to-day life during the testing. Although participants were allowed to complete the tests in any location, the vast majority completed them in their home environment, in part because data collection took place during the COVID-19 pandemic between November 2020 and April 2021. This may have reduced environmental influences to some extent. Nevertheless, a nontrivial number of participants were interrupted during the test or engaged in simultaneous activities that could be distracting, especially watching television or listening to music. Furthermore, our regression analyses showed that many of these environmental factors significantly affected the cognitive test scores. To evaluate the magnitude of these effects, apart from their statistical significance, it is useful to view them in the context of the corresponding age effects on cognitive scores. For the Figure Identification test, being interrupted was associated with a reduction of 0.89 figures correctly solved within the time limit (Table 5), which is approximately the same amount as the reduction in the Figure Identification test scores that would be expected for 3 years increase in age (Table 7). Similarly, for the SGST, being interrupted reduced participants' performance by approximately 0.08 to 0.15 seconds on average (Table 4), which corresponds with a performance reduction that would be expected for approximately 4 to 9 years increase in age (Table 6). This suggests that environmental distractions may have a nontrivial yet modest biasing impact when using observed cognitive scores in population-based research.

These environmental distractions occurred despite our instructing respondents to set aside uninterrupted time to complete the cognitive tasks. We developed an expanded warning about potential interruptions that concludes by requiring the participant to respond affirmatively that now is a good time to complete the tasks. We recommend that this approach be incorporated by other researchers using remote testing.

Our study had several limitations that should be considered. First, a nontrivial number of UAS panelists were not scored as they did not meet the accuracy threshold. Further investigation is important to understand the reasons (eg, to what extent this was because of inattentiveness vs the participants' lack of capability to complete the tasks) and the extent to which this introduced systematic bias in this study. Second, we developed tests in a probability panel, which mitigates the digital divide known to be associated with socioeconomic status by providing internet-connected devices to those who need them. However, it is possible that participants with low computer skills and poorer cognitive functioning were less likely to participate in this study. Third, our sample was predominantly English-speaking; the very small proportion of Spanish-speaking participants did not allow meaningful analyses or comparisons across language subgroups. Although the cognitive tests studied in this paper are less language dependent than many other commonly used neurocognitive tests, caution should be taken

XSL•FO
RenderX

when generalizing the study findings to specific subpopulations. Fourth, although the comparison of scores across device types was the primary objective of this study, there are many fine-grained differences within each of the device types (eg, screen or display size, keyboard, and touch screen functionality) that we did not examine, which could affect cognitive test scores. Fifth, although our sample included a group of older adults aged ≥75 years, the sample sizes were relatively small (311/6129, 5.07% for SGST; 419/6717, 6.24% for the Figure Identification test). Findings specific to this age group, such as the norms on cognitive scores, should be validated in larger samples in the future. Finally, our results suggest a small but unignorable impact of environmental distractions on test performance, which is an inherent problem for self-administered tests in general. In part, such inattentiveness may also lead to some participants failing to meet the accuracy threshold. Future studies are indispensable to detangle inattention from incapacity and explore ways of improving attention and reducing distractions.

Conclusions

Keeping these caveats in mind, we conclude that our rigorously developed cognitive measures are not unduly biased by the relative lack of standardization associated with web-based cognitive testing environments. The degree of error introduced by variations in devices and environments does not undermine the sensitivity of the measures used to detect group differences for research purposes. At the same time, we caution that the errors may be substantial enough to impede the accuracy of clinical decisions for individuals.

Our normative data, as presented, are suitable for interpreting SGST and the Figure Identification test results from future studies of English-speaking speaking US adult populations. To date, very few studies have provided normative data for web-based self-administered cognitive tests, and the quality of the norms provided here benefits from sampling weights developed within an existing probability-based sample and from larger sample sizes compared with previously reported web-based cognitive test data [1,3,4,12]. However, we also note that the samples used here are smaller than those used for validating and norming psychological tests in other areas such as quality of life research [45]; therefore, these numbers should be used with caution until more experience with these tests in web surveys has been gained.

Acknowledgments

This study was supported by grants from the National Institute on Aging of the National Institutes of Health (U01 AG054580 and R01 AG068190) and the Social Security Administration. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors.

The authors would like to thank Dr Arie Kapteyn, Principal Investigator of grant U01 AG054580, for his leadership in the Understanding America Study, consultation on this project, and feedback on the manuscript. The authors would also like to thank Dr Margie E Lachman for her consultation on the web adaptation of the Stop and Go Switching Task and Dr Boo Johansson for his consultation on the web adaptation of the Figure Identification test.

Authors' Contributions

MG and JED conceptualized the project, MG led the project, and YL and SS analyzed the data and prepared the first draft. The authorships of YL and SS are ordered alphabetically. BO programmed the web-based cognitive tests and device-detection algorithms. All authors took part in the discussion on study design, data collection, methodology, and results; contributed to the writing and revision of the manuscript; and approved the submission of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Demographic characteristics of the participants who completed the tests on both device types and met 70% correctness (included in the within-person comparison of test scores) and who owned both device types but did not meet these criteria (excluded from the within-person comparison).
[DOCX File , 24 KB - jmir_v24i5e34347_app1.docx]

Multimedia Appendix 2
Within-person comparison of cognitive scores across device types.
[DOCX File , 20 KB - jmir_v24i5e34347_app2.docx]

References

1. Tsoy E, Zygouris S, Possin KL. Current state of self-administered brief computerized cognitive assessments for detection of cognitive disorders in older adults: a systematic review. J Prev Alzheimers Dis 2021;8(3):267-276 [FREE Full text] [doi: 10.14283/jpad.2021.11] [Medline: 34101783]

XSL•FO
RenderX

2. Bissig D, Kaye J, Erten-Lyons D. Validation of SATURN, a free, electronic, self-administered cognitive screening test. Alzheimers Dement (N Y) 2020 Dec 29;6(1):e12116 [FREE Full text] [doi: 10.1002/trc2.12116] [Medline: 33392382]
3. Feenstra HE, Vermeulen IE, Murre JM, Schagen SB. Online self-administered cognitive testing using the Amsterdam cognition scan: establishing psychometric properties and normative data. J Med Internet Res 2018 May 30;20(5):e192 [FREE Full text] [doi: 10.2196/jmir.9298] [Medline: 29848469]
4. Ruano L, Sousa A, Severo M, Alves I, Colunas M, Barreto R, et al. Development of a self-administered web-based test for longitudinal cognitive assessment. Sci Rep 2016 Jan 08;6:19114 [FREE Full text] [doi: 10.1038/srep19114] [Medline: 26743329]
5. Schagen SB, Klein M, Reijneveld JC, Brain E, Deprez S, Joly F, et al. Monitoring and optimising cognitive function in cancer patients: present knowledge and future directions. Eur J Cancer Supplements 2014 Jun;12(1):29-40 [FREE Full text] [doi: 10.1016/j.ejcsup.2014.03.003] [Medline: 26217164]
6. Finkel D, Reynolds CA, McArdle JJ, Gatz M, Pedersen NL. Latent growth curve analyses of accelerating decline in cognitive abilities in late adulthood. Dev Psychol 2003 May;39(3):535-550. [doi: 10.1037/0012-1649.39.3.535] [Medline: 12760521]
7. Lachman ME, Agrigoroaei S, Tun PA, Weaver SL. Monitoring cognitive functioning: psychometric properties of the brief test of adult cognition by telephone. Assessment 2014 Aug;21(4):404-417 [FREE Full text] [doi: 10.1177/1073191113508807] [Medline: 24322011]
8. Langa KM, Llewellyn DJ, Lang IA, Weir DR, Wallace RB, Kabeto MU, et al. Cognitive health among older adults in the United States and in England. BMC Geriatr 2009 Jun 25;9:23 [FREE Full text] [doi: 10.1186/1471-2318-9-23] [Medline: 19555494]
9. Angrisani M, Finley B, Kapteyn A. Can Internet match high-quality traditional surveys? Comparing the Health and Retirement Study and its online version. In: Huynh KP, Jacho-chávez DT, Tripathi G, editors. The Econometrics of Complex Survey Data: Theory and Applications. Bingley, UK: Emerald Publishing Ltd; 2019.
10. Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, et al. Cognition assessment using the NIH toolbox. Neurology 2013 Mar 11;80(11 Supplement 3):S54-S64. [doi: 10.1212/wnl.0b013e3182872ded]
11. Birnbaum MH. Human research and data collection via the internet. Annu Rev Psychol 2004 Feb 01;55(1):803-832. [doi: 10.1146/annurev.psych.55.090902.141601] [Medline: 14744235]
12. Feenstra HE, Murre JM, Vermeulen IE, Kieffer JM, Schagen SB. Reliability and validity of a self-administered tool for online neuropsychological testing: the Amsterdam Cognition Scan. J Clin Exp Neuropsychol 2018 Apr;40(3):253-273. [doi: 10.1080/13803395.2017.1339017] [Medline: 28671504]
13. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized neuropsychological assessment devices: joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012 May;27(3):362-373 [FREE Full text] [doi: 10.1093/arclin/acs027] [Medline: 22382386]
14. Parsey CM, Schmitter-Edgecombe M. Applications of technology in neuropsychological assessment. Clin Neuropsychol 2013 Nov;27(8):1328-1361 [FREE Full text] [doi: 10.1080/13854046.2013.834971] [Medline: 24041037]
15. Passell E, Strong RW, Rutter LA, Kim H, Scheuer L, Martini P, et al. Cognitive test scores vary with choice of personal digital device. Behav Res Methods 2021 Dec;53(6):2544-2557. [doi: 10.3758/s13428-021-01597-3] [Medline: 33954913]
16. Cromer JA, Harel BT, Yu K, Valadka JS, Brunwin JW, Crawford CD, et al. Comparison of cognitive performance on the Cogstate brief battery when taken in-clinic, in-group, and unsupervised. Clin Neuropsychol 2015;29(4):542-558. [doi: 10.1080/13854046.2015.1054437] [Medline: 26165425]
17. Diamond A. Executive functions. Annu Rev Psychol 2013 Jan 03;64(1):135-168. [doi: 10.1146/annurev-psych-113011-143750]
18. Killgore WD. Effects of sleep deprivation on cognition. Prog Brain Res 2010;185:105-129. [doi: 10.1016/B978-0-444-53702-7.00007-5] [Medline: 21075236]
19. Korten NC, Comijs HC, Penninx BW, Deeg DJ. Perceived stress and cognitive function in older adults: which aspect of perceived stress is important? Int J Geriatr Psychiatry 2017 Apr;32(4):439-445. [doi: 10.1002/gps.4486] [Medline: 27059116]
20. Vance DE. Speed of processing in older adults: a cognitive overview for nursing. J Neurosci Nurs 2009 Dec;41(6):290-297. [doi: 10.1097/jnn.0b013e3181b6beda] [Medline: 19998680]
21. Tennstedt SL, Unverzagt FW. The ACTIVE study: study overview and major findings. J Aging Health 2013 Dec 02;25(8 Suppl):3S-20S [FREE Full text] [doi: 10.1177/0898264313518133] [Medline: 24385638]
22. Hughes ML, Agrigoroaei S, Jeon M, Bruzzese M, Lachman ME. Change in cognitive performance from midlife into old age: findings from the Midlife in the United States (MIDUS) study. J Int Neuropsychol Soc 2018 Jul 18;24(8):805-820. [doi: 10.1017/s1355617718000425]
23. Finkel D, Reynolds CA, McArdle JJ, Pedersen NL. Age changes in processing speed as a leading indicator of cognitive aging. Psychol Aging 2007 Sep;22(3):558-568. [doi: 10.1037/0882-7974.22.3.558] [Medline: 17874954]
24. Welcome to the understanding America study. USC Dornsife Center for Economic and Social Research. URL: https://uasdata.usc.edu/ [accessed 2022-04-14]
25. Alattar L, Messel M, Rogofsky D. An introduction to the understanding America Study internet panel. Soc Secur Bulletin 2018 May;78(2):13-28. [doi: 10.18356/f1091735-en]

XSL•FO
RenderX

26. A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples. In: Online Panel Research: A Data Quality Perspective. Brentwood, CA: John Wiley & Sons; 2014.
27. Yeager DS, Krosnick JA, Chang L, Javitz HS, Levendusky MS, Simpser A, et al. Comparing the accuracy of RDD telephone surveys and internet surveys conducted with probability and non-probability samples. Public Opin Q 2011 Oct 5;75(4):709-747. [doi: 10.1093/poq/nfr020]
28. Couper MP, Kapteyn A, Schonlau M, Winter J. Noncoverage and nonresponse in an internet survey. Social Sci Res 2007 Mar;36(1):131-148. [doi: 10.1016/j.ssresearch.2005.10.002]
29. Tun PA, Lachman ME. Age differences in reaction time and attention in a national telephone sample of adults: education, sex, and task complexity matter. Dev Psychol 2008 Sep;44(5):1421-1429 [FREE Full text] [doi: 10.1037/a0012845] [Medline: 18793073]
30. Thurstone LL. Some primary abilities in visual thinking. Proc Am Philosophical Soc 1950;94(6):517-521.
31. Dureman I, Kebbon L, Österberg E. Manual Till DS-Batteriet. Stockholm: Psykologiförlaget AB; 1971.
32. Berg S. Psychological functioning in 70- and 75-year-old people. A study in an industrialized city. Acta Psychiatr Scand Suppl 1980;288:1-47. [Medline: 6935934]
33. Welcome to NuBis. USC Dornsife Center for Economic and Social Research. URL: https://cesr.usc.edu/nubis/ [accessed 2022-04-14]
34. serbanghita/Mobile-Detect. URL: https://github.com/serbanghita/Mobile-Detect [accessed 2022-04-14]
35. Kyllonen PC, Zu J. Use of response time for measuring cognitive ability. J Intell 2016 Nov 01;4(4):14. [doi: 10.3390/jintelligence4040014]
36. Mead AD, Drasgow F. Equivalence of computerized and paper-and-pencil cognitive ability tests: a meta-analysis. Psychol Bull 1993 Nov;114(3):449-458. [doi: 10.1037/0033-2909.114.3.449]
37. Schneider S, May M, Stone AA. Careless responding in internet-based quality of life assessments. Qual Life Res 2018 Apr;27(4):1077-1088 [FREE Full text] [doi: 10.1007/s11136-017-1767-2] [Medline: 29248996]
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York, USA: Routledge; Jul 01, 1988.
39. Understanding America study: weights. USC Dornsife Center for Economic and Social Research. URL: https://uasdata.usc.edu/page/Weights [accessed 2022-04-14]
40. Cernich AN, Brennana DM, Barker LM, Bleiberg J. Sources of error in computerized neuropsychological assessment. Arch Clin Neuropsychol 2007 Feb;22 Suppl 1:S39-S48. [doi: 10.1016/j.acn.2006.10.004] [Medline: 17097851]
41. Nosek BA, Banaji MR, Greenwald AG. E-research: ethics, security, design, and control in psychological research on the internet. J Soc Issues 2002 Jan;58(1):161-176. [doi: 10.1111/1540-4560.00254]
42. Salthouse TA. When does age-related cognitive decline begin? Neurobiol Aging 2009 Apr;30(4):507-514 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2008.09.023] [Medline: 19231028]
43. Finkel D, Pedersen NL. Processing speed and longitudinal trajectories of change for cognitive abilities: the Swedish adoption/twin study of aging. Aging Neuropsychol Cognit 2010 Feb 02;11(2-3):325-345. [doi: 10.1080/13825580490511152]
44. Raz S, Bar-Haim Y, Sadeh A, Dan O. Reliability and validity of the online continuous performance test among young adults. Assessment 2014 Feb;21(1):108-118. [doi: 10.1177/1073191112443409] [Medline: 22517923]
45. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]

Abbreviations

SGST: Stop and Go Switching Task
UAS: Understanding America Study

Edited by R Kukafka; submitted 18.10.21; peer-reviewed by ZZ Weerts, I Niculescu; comments to author 27.12.21; revised version received 14.02.22; accepted 04.04.22; published 09.05.22.

Please cite as:
Liu Y, Schneider S, Orriens B, Meijer E, Darling JE, Gutsche T, Gatz M
Self-administered Web-Based Tests of Executive Functioning and Perceptual Speed: Measurement Development Study With a Large Probability-Based Survey Panel
J Med Internet Res 2022;24(5):e34347
URL: https://www.jmir.org/2022/5/e34347
doi:10.2196/34347
PMID:35532966

XSL•FO
RenderX

©Ying Liu, Stefan Schneider, Bart Orriens, Erik Meijer, Jill E Darling, Tania Gutsche, Margaret Gatz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.05.2022. 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.

XSL•FO
RenderX

Original Paper

Access to and Use of Internet and Social Media by Low-Morbidity Stroke Survivors Participating in a National Web-Based Secondary Stroke Prevention Trial: Cross-sectional Survey

Brigid Clancy[1,2], BHlth; Billie Bonevski[3], PhD; Coralie English[1], PhD; Amanda L Baker[1,2], PhD; Alyna Turner[1,4], PhD; Parker Magin[1], PhD; Michael Pollack[1,5], MBBS; Robin Callister[1,2], PhD; Ashleigh Guillaumier[1,2], PhD

[1]The University of Newcastle, Callaghan, Australia

[2]Hunter Medical Research Institute, New Lambton Heights, Australia

[3]Flinders University, Bedford Park, Australia

[4]Deakin University, Geelong, Australia

[5]Hunter New England Local Health District, New Lambton Heights, Australia

Corresponding Author:
Brigid Clancy, BHlth
The University of Newcastle
University Drive
Callaghan, 2308
Australia
Phone: 61 0432599892
Email: brigid.clancy@newcastle.edu.au

Abstract

Background: eHealth applications for stroke are a growing area of research that has yielded promising results. However, little is known about how stroke survivors engage with the internet, social media, and other digital technologies on a day-to-day basis.

Objective: This study had three main objectives: to describe the type, frequency, and purpose of technology use among a cohort of low-morbidity stroke survivors; to investigate associations between social media use and participant factors, including sociodemographics, physical function, and independence in activities of daily living; and to investigate associations between stroke-related health risk factors and the use of the internet to search for health and medical information.

Methods: This study is a secondary analysis of data obtained during a national randomized controlled trial—*Prevent 2nd Stroke*. The participants were stroke survivors recruited from 2 Australian stroke registries who completed 2 telephone-administered surveys to collect data on demographics and stroke characteristics; health risk factors (diet quality, physical activity, blood pressure medication, alcohol intake, anxiety and depression, and smoking status); physical functioning; independence in activities of daily living; and questions about what technology they had access to, how often they used it, and for what purposes. Participants were eligible if they had no more than a moderate level of disability (modified Rankin score ≤3) and had access to the internet. Multivariable logistic regression was used to assess the associations between social media use and sociodemographics, physical function, and independence in activities of daily living as well as associations between stroke-related health risk factors and the use of the internet to search for health and medical information.

Results: Data from 354 participants were included in the analysis. Approximately 79.1% (280/354) of participants used the internet at least daily, 40.8% (118/289) accessed social media on their phone or tablet daily, and 46.4% (134/289) looked up health and medical information at least monthly. Women were 2.7 times more likely to use social media (adjusted odds ratio 2.65, 95% CI 1.51-4.72), and people aged >75 years were significantly less likely to use social media compared with those aged <55 years (adjusted odds ratio 0.17, 95% CI 0.07-0.44). Health risk factors were not found to be associated with searching for health- or medical-related information.

Conclusions: The internet appears to be a viable platform to engage with stroke survivors who may not be high-morbidity to conduct research and provide information and health interventions. This is important given that they are at high risk of recurrent stroke regardless of their level of disability. Exploring the technology use behaviors and the possibility of eHealth among survivors who experience higher levels of morbidity or disability because of their stroke is an area of research that warrants further study.

XSL•FO
RenderX

(J Med Internet Res 2022;24(5):e33291) doi:10.2196/33291

KEYWORDS

stroke; stroke survivor; recurrent stroke; digital health; social media; internet use; eHealth; information-seeking behavior; web-based; mobile phone

Introduction

Background

Stroke is a leading cause of death and disability around the world [1]. Recurrent strokes are estimated to account for approximately one-quarter of all stroke incidents [2]. Data from the United States [3], United Kingdom [4], and Australia [5] suggest that stroke survivors do not receive adequate support and information on secondary stroke prevention following acute treatment. There is a need to develop better ways of reaching stroke survivors to involve them in health research as well as to deliver support and health information. Technology may be a part of this solution.

eHealth refers to health services or information that are delivered or enhanced through the use of the internet and related technologies [6]. eHealth has many potential applications in stroke prevention and care. Telerehabilitation has been one of the main areas of eHealth application in stroke and has been identified in multiple systematic reviews as a potential alternative or adjunct to usual rehabilitation [7-9], especially in the context of COVID-19 [10]. eHealth interventions aimed at reducing cardiovascular risk factors have shown promising results in glycemic control, achieving smoking cessation, diet and weight management, and increasing physical activity, and are generally acceptable and feasible [11-14]. A recent systematic review found that information and communication technology interventions aimed at stroke survivors and their carers are likely to provide some benefit, although the heterogeneity in study design and outcomes measured makes it difficult to draw nuanced conclusions [15].

Digital tools used in research studies, such as social media, data mining, email, and SMS text messaging, may be effective ways to engage stroke survivors in research. Digital tools are increasing in popularity for recruiting and retaining participants in randomized controlled trials, with the number of published studies doubling in the past decade [16]. Social media is a type of technology that is increasingly being used as a recruitment tool [17]. Little is known about how stroke survivors, who may be experiencing cognitive or physical disabilities, engage with social media and other technologies. Limitations on typing as well as a tendency for others' attention to focus on a person's disability have been associated with negative experiences for people with physical disabilities using social media [18].

The percentage of Australians using the internet to access health services has more than doubled from 22% in 2014-2015 to 46% in 2016-2017 [19]. Approximately 40% of Australians aged >55 years have accessed the internet for this purpose [19]. A study from the United States found that 57% of patients with acute coronary syndrome who had accessed the internet in the past 4 weeks reported web-based health information seeking [20]. However, there are no data available on web-based health and medical information seeking in people who have experienced stroke. A previous study on the web-based health information–seeking behaviors of people with chronic disease found that the third most common thing people searched for was lifestyle information such as diet and exercise [21]. It is possible that people who have experienced stroke are also looking to the internet for lifestyle information about health risk factors related to recurrent stroke, such as physical activity, diet, alcohol use, smoking, anxiety, and depression [22-24].

Despite the increasing popularity of eHealth use in this group, little is known about how stroke survivors engage with technology. A recent cross-sectional survey of stroke survivors (n=248) and carers (n=127) in the United States found that 81% of stroke survivors and 97% of carers had internet access and that smartphones were the most common device used to access the internet [25]. Beyond this, although information and communication technology interventions aimed at stroke survivors and their carers are likely to provide benefits, there is little information available about how stroke survivors use technology in their day-to-day lives [15]. A better understanding of this general use would assist in understanding how many people with lived experience of stroke have the potential to engage with eHealth programs or research.

Objectives

This study examined how a sample of Australian survivors of stroke who participated in a web-based secondary stroke prevention trial used technology. This study had three main objectives: (1) to describe the type, frequency, and purpose of technology use among a cohort of survivors of stroke; (2) to investigate associations between social media use and participant factors, including sociodemographics, physical function, and independence in activities of daily living; and (3) to investigate associations between stroke-related health risk factors (physical activity, diet, alcohol intake, smoking, and psychological distress) and using the internet to search for health and medical information.

Methods

Study Design

This study is a secondary analysis of cross-sectional data obtained during a national randomized controlled trial—*Prevent 2nd Stroke* [26,27]. Details regarding the study design and methods have been published elsewhere [26]. In short, trial participants (N=399) were recruited from the Australia-wide Australian Stroke Clinical Registry (AuSCR) and the Hunter Stroke Research Volunteer Registry (HSRVR), which recruits stroke survivors from the Hunter region of New South Wales.

The baseline survey was administered via a computer-assisted telephone interview and included questions about participant demographics, stroke characteristics, health risk behaviors, and

XSL•FO
RenderX

mood. The participants were then randomly allocated (1:1) to the intervention group, who received 12 weeks of access to the *Prevent 2nd Stroke* web-based program, or to the control group, who received internet addresses of readily available, generic web-based health programs designed for the general population. Six months after the initial baseline survey, a follow-up computer-assisted telephone interview survey was conducted (n=356), which included questions about the use of computers, tablets, phones, and the internet.

Ethics Approval

Ethical approval was obtained through the University of Newcastle Human Research Ethics Committee (H-2017-0051).

Participants

The AuSCR and HSRVR each used the eligibility criteria to screen their registrant databases and sent invitation packs on behalf of the study team to potentially eligible individuals. Individuals were eligible to take part in the study if they were aged ≥18 years, were part of the AuSCR or HSRVR, had had their most recent stroke in the previous 6-36 months, were fluent in English, and had access to the internet via a home device (eg, computer, tablet, or smartphone) or were willing to use public internet services (eg, public library). The participants were required to be able to walk without assistance and have no more than a moderate disability and so were ineligible if they scored ≥4 on the modified Rankin Scale [28].

Measures

Some measures were taken during the baseline survey either because they were static measures or to reduce the chance of health risk factor measures being affected by the intervention. The following measures were taken at the baseline survey.

Demographic and Stroke Characteristics

Age (<55, 55-64, 65-74, and >74 years), sex (male or female), weekly personal gross income (low: <Aus $399 [US $291.72], mid: Aus $400-$999 [US $292.45-$730.39], and high: ≥Aus $1000 [US $731.12]), whether they identify as Aboriginal or Torres Strait Islander or both, country of birth (Australia or other), stroke type (stroke or transient ischemic attack), and yes or no to whether it was their first stroke event.

Diet Quality

The Australian Recommended Food Score questionnaire was administered to assess usual diet quality, and scores were categorized as *needs work* (<33), *getting there* (33-38), *excellent* (39-46), or *outstanding* (≥47) [29].

Physical Activity

The Godin Leisure-Time Exercise Questionnaire was used to assess physical activity levels. The participants' responses were scored as *active* (≥24), *moderately active* (14-23), or *sedentary* (0-13) [30].

Blood Pressure Medication

Respondents were asked *Are you on blood pressure medications?* with response options *yes*, *no*, and *don't know*.

Alcohol Intake

The Alcohol Use Disorders Identification Test–Consumption was used to assess alcohol intake, and respondents with a score of ≥3 in women or ≥4 in men were considered to be drinking at potentially risky levels [31].

Psychological Distress, Anxiety, and Depression

The Patient Health Questionnaire–4 is an ultrabrief screening scale used as a measure of psychological distress and to screen for anxiety and depression [32]. Psychological distress was measured using all 4 questions and a score range of 0-12, with higher scores indicating greater psychological distress. Anxiety was measured using items 1 and 2, and depression was measured using items 3 and 4. A score of ≥3 for either pair of items was considered positive for anxiety or depression, respectively.

Smoking Status

Respondents were asked *do you currently smoke tobacco products?* with response options being *yes, daily*; *yes, once a week*; *yes, less than once a week*; and *no* [33]. This was used to determine the current smoking status of the participants.

Physical Functioning and Independent Living

The Barthel Index was used to assess physical functioning [34,35], where a score of 100 was considered *independent*, a score of 91-99 was considered *slight dependence*, and a score of 61-90 was considered *moderate dependence*. The Instrumental Activities of Daily Living was used to assess independent living ability [36], where a score of 8 was considered *independent living*, a score of 7 was considered *mostly independent living*, and a score of 0 to 6 was considered *requiring assistance*.

The technology-related measures were taken during the follow-up survey as this was the only stage at which they were asked. The potential effects of the intervention on these responses are discussed in the *Data Analysis* section. The participants were asked about what technology they had access to (including if it was internet-enabled); frequency of internet use; and how often they used their mobile phone or tablet for different purposes such as sending and receiving calls and SMS text messages, social media use, and accessing health- and medical-related information. These measures were adapted from the Australian Rural Mental Health Study [37] and the Pew Internet & American Life Poll 2012 Health Tracking Survey [38]. The survey questions used are presented in Multimedia Appendix 1.

Data Analysis

Descriptive statistics were used to describe the participants' access to technology, frequency of internet use, and purpose and frequency of use of a mobile phone or tablet. This included frequencies with percentages of nonmissing observations for categorical variables and mean with SD values for continuous variables.

Logistic regression modeling was used to identify associations between social media use and stroke survivor demographics, psychological distress, physical functioning, and independent living. Logistic regression modeling was also used to identify

XSL•FO
RenderX

whether the presence of health risk factors was associated with the use of a mobile phone or tablet to look for health- or medical-related information. Adjusted odds ratios (AORs) with 95% CIs and *P* values are provided. $P<.05$ was considered statistically significant.

The items regarding technology use were asked after the sample had participated in a trial encouraging technology use. To assess whether this affected the results (ie, overestimating typical use), logistic regression modeling was used to identify associations between group allocation and answers to the technology-related questions. No significant difference between groups in terms of how they used their mobile phone or tablet was found. There was a significant ($P=.02$) association between allocation and frequent use of internet (ie, once a week or more), with frequent internet use being 3.1 (95% CI 1.2-8.0) times more likely for people allocated to the intervention group than to the control group. However, because of the low number of participants who were infrequent internet users (25/354, 7.1%) and the fact that it did not appear to affect the responses used in the other logistic regressions on social media use and health and medical information, we determined that it was more appropriate to retain the larger sample size for this study than to exclude those in the intervention group.

Statistical analyses were programmed using R (version 4.0.2; R Foundation for Statistical Computing) [39].

Results

Participant Characteristics

Of the original 399 trial participants, 2 (0.5%) did not fully complete the follow-up survey and missed the technology-related questions, 38 (9.5%) were unable to be contacted for the follow-up survey, and 5 (1.3%) withdrew from the trial for personal reasons relating to their own health or care for others. In total, 354 participants completed the follow-up survey (333/354, 94.1% from the AuSCR and 21/354, 5.9% from the HSRVR). Most participants were male (231/354, 65.3%), and their average age was 68 (SD 12) years. Most participants were born in Australia (272/354, 76.8%). Table 1 shows the characteristics of the 354 participants who completed the follow-up survey.

XSL•FO
RenderX

Table 1. Sociodemographic and health risk factor characteristics of the participants (N=354).

Characteristic	Values[a]
Age (years), mean (SD)	68 (12)
Age categories (years), n (%)	
<55	41 (11.6)
55-64	71 (20.1)
65-74	136 (38.4)
>74	105 (29.7)
Men, n (%)	231 (65.2)
Income (Aus $; US $), n (%)	
Low (<399; <291.72)	97 (27.4)
Mid (400-999; 292.45-730.39)	139 (39.3)
High (≥1000; ≥731.12)	87 (24.6)
Do not know or did not answer	30 (8.5)
Aboriginal or Torres Strait Islander or both (yes), n (%)	1 (0.3)
Country of birth (Australia), n (%)	272 (76.8)
Stroke type, n (%)	
Stroke	216 (61)
TIA[b]	125 (35.3)
Do not know	13 (3.7)
History of previous stroke, n (%)	
No, first episode	328 (92.7)
Yes, had TIA before	15 (4.2)
Yes, had stroke before	5 (1.4)
Do not know	6 (1.7)
Diet quality (ARFS[c] score), n (%)	
Needs work (<33)	77 (21.8)
Getting there (33-38)	65 (18.4)
Excellent (39-46)	152 (42.9)
Outstanding (≥47)	101 (28.5)
Physical activity (GLTEQ[d] score), n (%)	
Sedentary (0-13)	116 (32.8)
Moderately active (14-25)	101 (28.5)
Active (≥24)	179 (50.6)
Blood pressure medication, n (%)	
No	111 (31.4)
Yes	275 (77.7)
Do not know	10 (2.8)
Potentially risky drinking (AUDIT-C[e]), n (%)	
No	207 (58.5)
Yes	189 (53.4)
Anxiety (PHQ-4[f] score), n (%)	
Low (0-2)	348 (98.3)

XSL•FO
RenderX

Characteristic	Values[a]
Anxiety (3-6)	46 (13)
Depression (PHQ-4 score), n (%)	
Low (0-2)	347 (98)
Depression (3-6)	46 (13)
Current smoker (yes), n (%)	16 (4.5)
Physical functioning (Barthel Index), n (%)	
Independent (100)	298 (84.2)
Slight dependence (91-99)	71 (20.1)
Moderate dependence (61-90)	24 (6.8)
Independent living (IADL[g] score), n (%)	
Independent living (8)	326 (92.1)
Mostly independent living (7)	39 (11)
Requiring assistance (0-6)	31 (8.8)

[a]Not all n values add up to 354 because of missing data.
[b]TIA: transient ischemic attack.
[c]ARFS: Australian Recommended Food Score.
[d]GLTEQ: Godin Leisure-Time Exercise Questionnaire.
[e]AUDIT-C: Alcohol Use Disorders Identification Test–Consumption.
[f]PHQ-4: Patient Health Questionnaire–4.
[g]IADL: Instrumental Activities of Daily Living.

Type, Frequency, and Purpose of Technology Use

Table 2 shows the participants' access to technology. Most participants (319/354, 90.1%) had access to a computer with internet access, and 73.4% (260/354) had access to a mobile phone with internet access.

Table 3 shows the frequency of internet use among the whole sample of participants, with most accessing the internet at least daily (280/354, 79.1%) and only 4.8% (17/354) answering that they did not use the internet.

Table 4 shows the purpose and frequency of use of the participants who answered *yes* to having access to a mobile phone or tablet with internet access in Table 2 (289/354, 81.6%). Half of these participants used their mobile phone or tablet to access social media networking sites at least weekly (149/289, 51.6%), whereas 46.4% (134/289) looked for health- or medical-related information at least monthly. Mobile phones and tablets were used daily for communication purposes, including phone calls (234/289, 81%), text messages (203/289, 70.2%), and emails (139/289, 48.1%).

Table 2. Access to technology at home or elsewhere (N=354).

Access to technology	Yes, n (%)
Computer with internet access	319 (90.1)
Mobile phone with internet access	260 (73.4)
Tablet device with internet access	209 (59)
Webcam	113 (31.9)
Mobile phone without internet access	73 (20.6)
Tablet device without internet access	13 (3.7)
Computer without internet access	5 (1.4)
None of the above	2 (0.6)

XSL•FO
RenderX

Table 3. Frequency of internet use, including email (N=354).

Frequency of internet use	Total, n (%)
Several times a day	174 (49.2)
Every day	106 (29.9)
Several times a week	33 (9.3)
Once a week	16 (4.5)
Once a month or less	7 (2)
I do not use the internet	17 (4.8)
Do not know	1 (0.3)

Table 4. Purpose and frequency of use of mobile phone or tablet (N=289).

	Never, n (%)	Less than monthly, n (%)	Monthly, n (%)	Weekly, n (%)	Daily, n (%)	Do not know, n (%)
Making or receiving phone calls	6 (2.1)	4 (1.4)	12 (4.2)	32 (11.1)	234 (81)	1 (0.3)
Sending or receiving SMS text messages	23 (8)	6 (2.1)	4 (1.4)	52 (18)	203 (70.2)	1 (0.3)
Accessing the internet	36 (12.5)	7 (2.4)	6 (2.1)	43 (14.9)	196 (67.8)	1 (0.3)
Using apps	55 (19)	12 (4.2)	14 (4.8)	36 (12.5)	169 (58.5)	3 (1)
Accessing social media networking sites (eg, Facebook, Twitter, or Instagram)	127 (43.9)	8 (2.8)	5 (1.7)	31 (10.7)	118 (40.8)	0 (0)
Sending or receiving emails	61 (21.1)	13 (4.5)	13 (4.5)	62 (21.5)	139 (48.1)	1 (0.3)
Taking pictures (photos)	40 (13.8)	30 (10.4)	66 (22.8)	95 (32.9)	58 (20.1)	0 (0)
Looking for health- or medical-related information	112 (38.8)	42 (14.5)	63 (21.8)	53 (18.3)	18 (6.2)	1 (0.3)
For entertainment (eg, music, watching videos, or Netflix)	138 (47.8)	15 (5.2)	20 (6.9)	45 (15.6)	70 (24.2)	1 (0.3)

Associations Between Participant Characteristics and Social Media Use

Table 5 shows the multivariable regression results for outcome use of phone or tablet to access social media networking sites. It includes only the participants who answered *yes* to having access to a mobile phone or tablet with internet access in Table 2 (289/354, 81.6%). After accounting for the other variables in the model, age and sex showed a statistically significant association with social media use, with younger persons and women more likely to use social media. Women were 2.7 times more likely to use social media compared with men (AOR 2.65, 95% CI 1.51-4.72). Being aged >55 years was associated with lower odds of using social media compared with those aged <55 years; however, only the highest age category (≥75 years) was significantly different from <55 years (AOR 0.17, 95% CI 0.07-0.44). None of the other variables assessed showed a statistically significant association.

XSL•FO
RenderX

Table 5. Multivariable regression results for outcome use of phone or tablet to access social media networking sites (N=289).

Variable	Use of social media via phone or tablet, n (%)		Multivariable	
	No	Yes	OR[a] (95% CI)	*P* value
Age categories (years)	135 (100)	153 (100)		.001
<55	10 (7.4)	31 (20.3)	__[b]	
55-64	29 (21.5)	35 (22.9)	0.46 (0.18-1.11)	
65-74	52 (38.5)	68 (44.4)	0.50 (0.21-1.15)	
>74	44 (32.6)	19 (12.4)	0.17 (0.07-0.44)	
Sex	135 (100)	154 (100)		.001
Male	100 (74.1)	83 (53.9)	—	
Female	35 (25.9)	71 (46.1)	2.65 (1.51-4.72)	
Income (Aus $; US $)	135 (100)	154 (100)		.83
Low (<399; <291.72)	38 (28.1)	44 (28.6)	—	
Mid (400-999; 292.45-730.39)	52 (38.5)	54 (35.1)	1.03 (0.54-1.96)	
High (>999; >730.39)	33 (24.4)	46 (29.9)	1.29 (0.61-2.74)	
Do not know or refused	12 (8.9)	10 (6.5)	0.80 (0.29-2.21)	
Psychological distress (PHQ-4[c] score)	134 (100)	153 (100)		.87
None (0-2)	100 (74.6)	116 (75.8)	—	
Mild (3-5)	24 (17.9)	24 (15.7)	0.84 (0.42-1.68)	
Moderate or severe (6-12)	10 (7.5)	13 (8.5)	1.04 (0.40-2.76)	
Physical functioning (Barthel Index)	135 (100)	154 (100)		.83
Independent (100)	104 (77)	120 (77.9)	—	
Slight dependence (91-99)	24 (17.8)	26 (16.9)	0.83 (0.40-1.71)	
Moderate dependence (61-90)	7 (5.2)	8 (5.2)	0.75 (0.21-2.68)	
Independent living (IADL[d] score)	135 (100)	154 (100)		.91
Independent living (8)	117 (86.7)	134 (87)	—	
Mostly independent living (7)	10 (7.4)	10 (6.5)	1.16 (0.42-3.21)	
Requiring assistance (0-6)	8 (5.9)	10 (6.5)	1.25 (0.39-4.17)	

[a]OR: odds ratio.

[b]There was no comparison in these cells.

[c]PHQ-4: Patient Health Questionnaire–4.

[d]IADL: Instrumental Activities of Daily Living.

Associations Between Health Risk Factors and Health Information Searching

In the multivariable regression conducted to determine whether the use of a phone or tablet to find medical information was related to the presence of health risk factors, no results were significant. None of the health risk factor variables were associated with searching for medical information on the web on a phone or tablet.

Discussion

Principal Findings

This study found that most participants owned a computer, mobile phone, or tablet with internet access and accessed the internet at least daily. Internet access was a requirement for participation; therefore, these are unsurprising results for this particular cohort. Daily activities on these devices for most of the sample included making or receiving phone calls, sending or receiving SMS text messages, accessing the internet, or using apps. Most also accessed social media at least weekly (149/289, 51.6%), and almost half of them accessed health- or medical-related information at least monthly (134/289, 46.4%).

We found that stroke survivors, including those in older age groups, frequently use the internet for a number of purposes. A previous US study of stroke survivors with a similar average age (64 years) reported that 81% had access to the internet, with most accessing it for >5 hours per week [25], whereas a Danish study of 100 people who were patients in a stroke unit found that 87% reported having access to an internet-enabled device

XSL•FO
RenderX

at home [40]. In 2016-2017 in Australia, 55% of the general population aged >65 years had used the internet in the last 3 months, an increase on the 46% from 2012 to 2013, and two-thirds had accessed it for health services [19]. This shows that, although internet use among older people may be lower than among the rest of the population, there is still a significant proportion of older people engaging with it, and this use is on the rise. The increasing rates of stroke among younger adults cannot be ignored when examining technology use among stroke survivors. Hospitalization rates for acute ischemic stroke have increased significantly between 2003 and 2012 among men (41.5%) and women (30%) aged between 35 and 44 years [41]. With the rate of internet access among this group at 96% [19], it is important to consider this younger age group in the development of internet-based interventions. It can be inferred that, with the increasing number of young people experiencing stroke, in combination with the increasing use of the internet by older age groups and the already significant use of the internet in the current stroke survivor population, digital literacy among those who have experienced stroke will only become more prevalent. As digital literacy increases and more health and medical information and interventional programs are available on the web, a targeted effort is needed to ensure that stroke survivors who are interested in these sorts of programs, including those who face communication difficulties, are not left behind.

Previous research has found that stroke-related information provision for people with lived experience of stroke is insufficient [42,43]. More than 50% of stroke survivors have self-reported an unmet need for stroke information [4]. In Australia, only 63% of patients in inpatient stroke rehabilitation services receive education about stroke, lifestyle management, secondary stroke prevention, and recovery [5]. With many stroke survivors having access to the internet, it is not unreasonable to assume that some are looking on the web to fill these information gaps. Approximately 46.4% (134/289) of our sample used their mobile phones or tablet devices to look up health or medical information at least once a month. This is congruent with previous research that found that 57% of patients with acute coronary syndrome who had accessed the internet in the past 4 weeks had used it for web-based health information seeking [20]. The quality and accessibility of the information available to stroke survivors on websites are mixed [44], and much of this information does not meet the recommended readability guidelines for stroke survivors [45]. Some people who have experienced stroke also turn to unregulated options such as web-based forums for information [46]. With many stroke survivors lacking adequate stroke-related information and turning to the internet for additional health and medical information, there is a need to ensure that the information available on the web is not only accurate and appropriate but also accessible for this population.

Although digital health presents opportunities for greater reach to promote health, it is not a one-size-fits-all solution. It is likely to be most suited to those with lower morbidity or disability, with people who are more greatly affected by stroke requiring more intensive resources [47]. Nonetheless, all stroke survivors are at higher risk of recurrent stroke than the general population [48], and all avenues should be explored to reduce their risk of recurrent stroke. This study suggests that there is an opportunity to use digital health applications to reduce the risk of recurrent stroke among survivors with low levels of disability. Further studies will be required to explore this application in people who experience greater morbidity and disability as a result of stroke.

Among our sample of low-morbidity stroke survivors, we investigated whether there was a relationship between the presence of health risk factors (diet quality, physical activity, blood pressure medication, alcohol intake, anxiety and depression, and smoking status) and using the internet to search for health-related information. However, we did not find any associations. This may be due to a number of reasons, such as the survey questions used not being specific enough; individuals not being aware of their risk or not seeing their behavior as problematic; or participants obtaining that type of information elsewhere, such as a managing general practitioner. However, most of this sample looked for health information on the web regardless of their health risk factors. This presents a clear opportunity to provide another mode of education and engagement around health risk factors and risk of recurrent stroke in a population that often misses out on receiving secondary stroke prevention education [4,5].

Social media platforms are also of increasing interest to health researchers [49]. They offer a low or no-cost means of observing and reaching both diverse and narrow audiences with the possibility of multidirectional communication [49]. Social media was identified as an effective recruitment tool in a randomized controlled trial for hypertension [50]. In the general Australian population, 51% of people aged >65 years who had accessed the internet in the last 3 months had accessed social media during this time [19]. We had similar findings within our sample, with just over half of the participants (149/289, 51.6%) accessing social media sites at least weekly, whereas 43.9% (127/289) did not access them at all. On the basis of our results, researchers looking to use social media to observe or access stroke survivor populations may face more difficulty accessing male stroke survivors and those aged ≥75 years. Social media may be a feasible platform for recruiting some groups of stroke survivors.

Strengths and Limitations

This study provides data from a large national sample of stroke survivors. The participants in this study were technology users by nature given their recruitment to a web-based secondary stroke prevention trial where the use of an internet-enabled device and email were required. They were also a relatively *well* cohort compared with the general stroke population, with high levels of independence and limited disability because of the eligibility requirement of a modified Rankin score of ≤3. Unfortunately, this means that the trial excluded many of those who are significantly affected by their stroke-related impairments and cannot be generalized to the Australian stroke survivor population as a whole. More research into the internet use behaviors of those with greater levels of disability and dependence, as well as how specific impairments may affect the use of the internet, is warranted. However, this study does provide an indication of the patterns of technology use among

XSL•FO
RenderX

more able stroke survivors who are already internet users and willing to participate in internet-based research. This population still requires support in managing their health risk factors, and this research is beneficial for better understanding how to access them and deliver appropriate care.

The questions assessing social media use, accessing health- and medical-related information, and other device-related activities were also only asked in the context of mobile phone and tablet use. The questions did not account for alternative modes of internet access such as a desktop or laptop computer, which may lead to an underrepresentation of stroke survivor engagement with these activities.

Conclusions

The internet may be a viable platform to engage with stroke survivors experiencing low levels of disability for health interventions, information, and research. This is important as all stroke survivors are at higher risk of stroke than the general population and require secondary stroke prevention support and education. Exploring the technology use behaviors and possibility of eHealth with people who experience greater levels of disability following stroke is an area of research that warrants further study.

Acknowledgments

This work was supported by a grant from the National Health and Medical Research Council (APP1125429). BC is supported by a Research Training Program scholarship from the University of Newcastle. AG is supported by a postdoctoral fellowship from the Heart Foundation. ALB is supported by a National Health and Medical Research Council Research Fellowship. The authors wish to acknowledge and thank all the participating stroke survivors. The authors also wish to thank Jack Faulkner from the Hunter Medical Research Institute for his significant contributions to and assistance with data analysis.

Authors' Contributions

BC contributed to conceptualization; methodology; data curation; data analysis; and writing, reviewing, and editing the manuscript. BB contributed to supervision, conceptualization, methodology, and review and editing of the manuscript. CE contributed to supervision and review and editing of the manuscript. ALB, PM, AT, MP, and RC contributed to conceptualization, methodology, and reviewing and editing the manuscript. AG contributed to supervision, conceptualization, survey development, methodology, and review and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Technology use survey questions.
[DOCX File , 15 KB - jmir_v24i5e33291_app1.docx]

References

1. Leading causes of death and disability: a visual summary of global and regional trends 2000-2019. World Health Organization. URL: https://www.who.int/data/stories/leading-causes-of-death-and-disability-2000-2019-a-visual-summary [accessed 2021-09-05]
2. Rothwell PM, Coull AJ, Giles MF, Howard SC, Silver LE, Bull LM, Oxford Vascular Study. Change in stroke incidence, mortality, case-fatality, severity, and risk factors in Oxfordshire, UK from 1981 to 2004 (Oxford Vascular Study). Lancet 2004 Jun 12;363(9425):1925-1933. [doi: 10.1016/S0140-6736(04)16405-2] [Medline: 15194251]
3. Klinedinst NJ, Dunbar SB, Clark PC. Stroke survivor and informal caregiver perceptions of poststroke depressive symptoms. J Neurosci Nurs 2012 Apr;44(2):72-81 [FREE Full text] [doi: 10.1097/JNN.0b013e3182477944] [Medline: 22367269]
4. McKevitt C, Fudge N, Redfern J, Sheldenkar A, Crichton S, Rudd AR, et al. Self-reported long-term needs after stroke. Stroke 2011 May;42(5):1398-1403. [doi: 10.1161/STROKEAHA.110.598839] [Medline: 21441153]
5. National Stroke Audit - Rehabilitation Services Report 2020. Stroke Foundation. 2020. URL: https://informme.org.au/media/drtlcbvp/rehab_strokeservicesreport_2020.pdf [accessed 2021-07-22]
6. Eysenbach G. What is e-health? J Med Internet Res 2001 Apr;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
7. Tchero H, Tabue Teguo M, Lannuzel A, Rusch E. Telerehabilitation for stroke survivors: systematic review and meta-analysis. J Med Internet Res 2018 Oct 26;20(10):e10867 [FREE Full text] [doi: 10.2196/10867] [Medline: 30368437]
8. Appleby E, Gill ST, Hayes LK, Walker TL, Walsh M, Kumar S. Effectiveness of telerehabilitation in the management of adults with stroke: a systematic review. PLoS One 2019 Nov 12;14(11):e0225150 [FREE Full text] [doi: 10.1371/journal.pone.0225150] [Medline: 31714924]

XSL•FO
RenderX

9. Laver KE, Adey-Wakeling Z, Crotty M, Lannin NA, George S, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020 Jan 31;1:CD010255 [FREE Full text] [doi: 10.1002/14651858.CD010255.pub3] [Medline: 32002991]
10. Chang MC, Boudier-Revéret M. Usefulness of telerehabilitation for stroke patients during the COVID-19 pandemic. Am J Phys Med Rehabil 2020 Jul;99(7):582 [FREE Full text] [doi: 10.1097/PHM.0000000000001468] [Medline: 32404640]
11. Liu S, Feng W, Chhatbar PY, Liu Y, Ji X, Ovbiagele B. Mobile health as a viable strategy to enhance stroke risk factor control: a systematic review and meta-analysis. J Neurol Sci 2017 Jul 15;378:140-145 [FREE Full text] [doi: 10.1016/j.jns.2017.04.050] [Medline: 28566151]
12. 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 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
13. Devi R, Singh SJ, Powell J, Fulton EA, Igbinedion E, Rees K. Internet-based interventions for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2015 Dec 22(12):CD009386. [doi: 10.1002/14651858.CD009386.pub2] [Medline: 26691216]
14. 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 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
15. Freund M, Carey M, Dilworth S, Waller A, Mansfield E, Rose A, et al. Effectiveness of information and communications technology interventions for stroke survivors and their support people: a systematic review. Disabil Rehabil 2021 Apr 27:1-16. [doi: 10.1080/09638288.2021.1913245] [Medline: 33905279]
16. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020 Jun 05;21(1):478 [FREE Full text] [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
17. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
18. Lee HE, Cho J. Social media use and well-being in people with physical disabilities: influence of SNS and online community uses on social support, depression, and psychological disposition. Health Commun 2019 Aug;34(9):1043-1052. [doi: 10.1080/10410236.2018.1455138] [Medline: 29652521]
19. Household use of information technology, Australia, 2016-17. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/industry/technology-and-innovation/household-use-information-technology/latest-release [accessed 2021-06-18]
20. Waring ME, McManus DD, Amante DJ, Darling CE, Kiefe CI. Online health information seeking by adults hospitalized for acute coronary syndromes: who looks for information, and who discusses it with healthcare providers? Patient Educ Couns 2018 Nov;101(11):1973-1981 [FREE Full text] [doi: 10.1016/j.pec.2018.06.016] [Medline: 30305253]
21. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014 Dec 02;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
22. O'Donnell MJ, Xavier D, Liu L, Zhang H, Chin SL, Rao-Melacini P, INTERSTROKE investigators. Global and regional effects of potentially modifiable risk factors associated with acute stroke in 32 countries (INTERSTROKE): a case-control study. Lancet 2016 Aug 20;388(10046):761-775. [doi: 10.1016/S0140-6736(10)60834-3] [Medline: 20561675]
23. Wu QF, Zhou AM, Han YP, Liu YM, Yang Y, Wang XM, et al. Poststroke depression and risk of recurrent stroke: a meta-analysis of prospective studies. Medicine (Baltimore) 2019 Oct;98(42):e17235 [FREE Full text] [doi: 10.1097/MD.0000000000017235] [Medline: 31626084]
24. Pérez-Piñar M, Ayerbe L, González E, Mathur R, Foguet-Boreu Q, Ayis S. Anxiety disorders and risk of stroke: a systematic review and meta-analysis. Eur Psychiatry 2017 Mar;41:102-108. [doi: 10.1016/j.eurpsy.2016.11.004] [Medline: 28135591]
25. Naqvi IA, Montiel TC, Bittar Y, Hunter N, Okpala M, Johnson C, et al. Internet access and usage among stroke survivors and their informal caregivers: cross-sectional study. JMIR Form Res 2021 Mar 08;5(3):e25123 [FREE Full text] [doi: 10.2196/25123] [Medline: 33683206]
26. Guillaumier A, McCrabb S, Spratt NJ, Pollack M, Baker AL, Magin P, et al. An online intervention for improving stroke survivors' health-related quality of life: study protocol for a randomised controlled trial. Trials 2019 Aug 09;20(1):491 [FREE Full text] [doi: 10.1186/s13063-019-3604-0] [Medline: 31399140]
27. Guillaumier A, Spratt NJ, Pollack M, Baker A, Magin P, Turner A, et al. Evaluation of an online intervention for improving stroke survivors' health-related quality of life: a randomised controlled trial. PLoS Med 2022 Apr 19;19(4):e1003966 [FREE Full text] [doi: 10.1371/journal.pmed.1003966] [Medline: 35439246]
28. Quinn TJ, Lees KR, Hardemark HG, Dawson J, Walters MR. Initial experience of a digital training resource for modified Rankin scale assessment in clinical trials. Stroke 2007 Aug;38(8):2257-2261. [doi: 10.1161/STROKEAHA.106.480723] [Medline: 17600236]

XSL•FO
RenderX

29. Collins CE, Burrows TL, Rollo ME, Boggess MM, Watson JF, Guest M, et al. The comparative validity and reproducibility of a diet quality index for adults: the Australian Recommended Food Score. Nutrients 2015 Jan 23;7(2):785-798 [FREE Full text] [doi: 10.3390/nu7020785] [Medline: 25625814]
30. Eisenmann JC, Milburn N, Jacobsen L, Moore SJ. Reliability and convergent validity of the Godin Leisure-Time Exercise Questionnaire in rural 5th-grade school-children. J Hum Mov Stud 2002 Jan;43(2):135-149. [doi: 10.1097/00005768-199706001-00009]
31. Bush K, Kivlahan DR, McDonell 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] [Medline: 9738608]
32. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
33. Mullins R, Borland R. Changing the way smoking is measured among Australian adults: a preliminary investigation of Victorian data. Quit Evaluation Studies. 1998. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.582.9632&rep=rep1&type=pdf [accessed 2021-06-15]
34. Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. Md State Med J 1965 Feb;14:61-65. [Medline: 14258950]
35. Middleton S, Levi C, Ward J, Grimshaw J, Griffiths R, D'Este C, et al. Fever, hyperglycaemia and swallowing dysfunction management in acute stroke: a cluster randomised controlled trial of knowledge transfer. Implement Sci 2009 Mar 16;4:16 [FREE Full text] [doi: 10.1186/1748-5908-4-16] [Medline: 19291323]
36. Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. J Am Geriatr Soc 1983 Dec;31(12):721-727. [doi: 10.1111/j.1532-5415.1983.tb03391.x] [Medline: 6418786]
37. Kelly BJ, Stain HJ, Coleman C, Perkins D, Fragar L, Fuller J, et al. Mental health and well-being within rural communities: the Australian rural mental health study. Aust J Rural Health 2010 Feb;18(1):16-24. [doi: 10.1111/j.1440-1584.2009.01118.x] [Medline: 20136810]
38. Pew Internet & American Life Project. 2012. URL: https://ropercenter.cornell.edu/ipoll/study/31096124 [accessed 2022-03-22]
39. R Foundation for Statistical Computing. URL: http://www.R-project.org/ [accessed 2022-03-22]
40. Ciccone NW, Dornonville de la Cour FL, Thorpe JR, Forchhammer BH, Maier A. Personal technology use amongst stroke patients: understanding the best platforms for the design of health interventions in treatment and rehabilitation. In: Proceedings of the Design Society: 23rd International Conference on Engineering Design. 2021 Jul 27 Presented at: ICED '21; August 16-20, 2021; Gothenburg, Sweden p. 2419-2428. [doi: 10.1017/pds.2021.503]
41. George MG, Tong X, Bowman BA. Prevalence of cardiovascular risk factors and strokes in younger adults. JAMA Neurol 2017 Jun 01;74(6):695-703 [FREE Full text] [doi: 10.1001/jamaneurol.2017.0020] [Medline: 28395017]
42. Chen T, Zhang B, Deng Y, Fan J, Zhang L, Song F. Long-term unmet needs after stroke: systematic review of evidence from survey studies. BMJ Open 2019 May 19;9(5):e028137 [FREE Full text] [doi: 10.1136/bmjopen-2018-028137] [Medline: 31110106]
43. Faiz KW, Labberton AS, Thommessen B, Rønning OM, Barra M. Stroke-related knowledge and lifestyle behavior among stroke survivors. J Stroke Cerebrovasc Dis 2019 Nov;28(11):104359. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104359] [Medline: 31495671]
44. Surman R, Bath PA. An assessment of the quality of information on stroke and speech and language difficulty web sites. J Inf Sci 2013 Jan 08;39(1):113-125. [doi: 10.1177/0165551512469775]
45. Sharma N, Tridimas A, Fitzsimmons PR. A readability assessment of online stroke information. J Stroke Cerebrovasc Dis 2014 Jul;23(6):1362-1367. [doi: 10.1016/j.jstrokecerebrovasdis.2013.11.017] [Medline: 24389374]
46. De Simoni A, Shanks A, Balasooriya-Smeekens C, Mant J. Stroke survivors and their families receive information and support on an individual basis from an online forum: descriptive analysis of a population of 2348 patients and qualitative study of a sample of participants. BMJ Open 2016 Apr 06;6(4):e010501 [FREE Full text] [doi: 10.1136/bmjopen-2015-010501] [Medline: 27053271]
47. Lekander I, Willers C, von Euler M, Lilja M, Sunnerhagen KS, Pessah-Rasmussen H, et al. Relationship between functional disability and costs one and two years post stroke. PLoS One 2017;12(4):e0174861 [FREE Full text] [doi: 10.1371/journal.pone.0174861] [Medline: 28384164]
48. Flach C, Muruet W, Wolfe CD, Bhalla A, Douiri A. Risk and secondary prevention of stroke recurrence: a population-base cohort study. Stroke 2020 Aug;51(8):2435-2444 [FREE Full text] [doi: 10.1161/STROKEAHA.120.028992] [Medline: 32646337]
49. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
50. Feyz L, Wang Y, Pathak A, Saxena M, Mahfoud F, Sanghvi K, et al. Using social media to recruit study participants for a randomized trial for hypertension. Eur Heart J 2020 Nov 30;1(1):71-74. [doi: 10.1093/ehjdh/ztaa010]

XSL•FO
RenderX

Abbreviations

AOR: adjusted odds ratio
AuSCR: Australian Stroke Clinical Registry
HSRVR: Hunter Stroke Research Volunteer Registry

Edited by R Kukafka; submitted 08.09.21; peer-reviewed by T Langford, M El Tantawi; comments to author 01.12.21; revised version received 17.02.22; accepted 23.02.22; published 30.05.22.

Please cite as:
Clancy B, Bonevski B, English C, Baker AL, Turner A, Magin P, Pollack M, Callister R, Guillaumier A
Access to and Use of Internet and Social Media by Low-Morbidity Stroke Survivors Participating in a National Web-Based Secondary Stroke Prevention Trial: Cross-sectional Survey
J Med Internet Res 2022;24(5):e33291
URL: https://www.jmir.org/2022/5/e33291
doi:10.2196/33291
PMID:35635754

©Brigid Clancy, Billie Bonevski, Coralie English, Amanda L Baker, Alyna Turner, Parker Magin, Michael Pollack, Robin Callister, Ashleigh Guillaumier. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.05.2022. 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.

XSL•FO
RenderX

Original Paper

Self-efficacy and Use of Digital Health Care and Social Welfare Services Among Incarcerated People: Cross-sectional Survey Study

Teemu Rantanen[1], DPhil; Eeva Järveläinen[1], MA; Teppo Leppälahti[1], MA

Laurea University of Applied Sciences, Vantaa, Finland

Corresponding Author:
Teemu Rantanen, DPhil
Laurea University of Applied Sciences
Ratatie 22
Vantaa, 01300
Finland
Phone: 358 408306149
Email: teemu.rantanen@laurea.fi

Abstract

Background: The digitization of health care and social welfare services creates many opportunities for the rehabilitation of incarcerated people and their preparation for release from prison. A range of digital platforms and technology solutions have been developed that offer multiple opportunities to handle private matters either by video conference, email, or some other digital format during imprisonment. However, incarcerated people have limited access to digital health care and social welfare services, and face challenges related to shortcomings in their digital skills and self-efficacy.

Objective: This article assessed the significance of incarcerated people's self-efficacy in terms of their sense of control over the use of digital health care and social welfare services.

Methods: A cross-sectional study was conducted using a questionnaire. Research data were collected from 11 prisons in different parts of Finland, and a total of 225 incarcerated people responded to the survey. Statistical analyses were conducted using the Pearson product-moment correlation coefficient, 2-tailed t test, linear regression analysis, and Hayes bootstrapping method.

Results: The results showed the significance of both general and internet-specific self-efficacy, which appear to be more important for the use of digital health care and social welfare services than factors related to a person's socioeconomic background or sentence. Age was negatively correlated with perceived control over the use of digital health care and social welfare services. Furthermore, the study emphasized the importance of support from family and friends, as well as prison employees.

Conclusions: The digitalization of prisons offers many opportunities, but special consideration should be given to how the digitization of health care and social welfare services responds to the needs of incarcerated people in terms of their integration into society and the prevention of recidivism. During imprisonment, attention should be paid to strengthening the digital skills of incarcerated people, with support provided by prison employees. In addition to providing guidance on the use of individual digital services, the study recommends strengthening the general digital skills of incarcerated people, as well as developing their life management skills.

(J Med Internet Res 2022;24(5):e36799) doi:10.2196/36799

KEYWORDS

digital inclusion; digital exclusion; digital divide; digitalization in prisons; incarcerated people

Introduction

Role of Digital Health Care and Social Welfare Services in Supporting a Desistance From Crime

Recent discussions on the rehabilitation of incarcerated people and their re-entry into society have emphasized the perspective of a desistance from crime [1-4]. Breaking away from crime is often a very challenging and long-lasting social process for a person with a history of crime, which can include single sudden events [5] or a series of events in a person's life [2,3]. It is crucial to ensure that timely and adequate support and rehabilitation services are made available during the prison term and in the release phase, and that support continues to outside

XSL•FO
RenderX

life [4,6,7]. In addition, while addressing the prevention of recidivism, attention should be paid to the social situation and its improvement [6,8].

Previous studies have argued that incarcerated people should have access to digital services already during their imprisonment, in order to facilitate their integration into society [7,9]. Consistent with this, a range of digital platforms and technology solutions have been developed that offer opportunities to handle private matters either by video conference, email, or some other digital format during imprisonment, with the aim of supporting incarcerated people's re-entry into society [10-12]. Digitalization in prison makes it possible for incarcerated people to use digital health care and social welfare services, participate in distance learning or rehabilitation, and communicate with family or different authorities [13]. Reisdorf and Rikard [7] have highlighted that both offline and online rehabilitation measures should be targeted during imprisonment and in the re-entry phase, in order to strengthen returnee agency in a digital society. Furthermore, Ogbonnaya-Ogburu et al [14] emphasized the importance of digital literacy in relation to job-searching skills, which also makes it easier for incarcerated people to re-enter society.

Digital and social inclusion presupposes access to the internet, adequate digital skills, a positive attitude toward online services [15,16], and a reasonable degree of prison employee trust in incarcerated people [17]. However, during imprisonment, access to digital health care and social welfare services is limited, and the use of electronic services is characterized by a lack of trust, the use of control, and a culture that emphasizes formal security [18]. Furthermore, incarcerated people often lack digital skills and display attitudinal problems associated with the use of online services [14,19], although they are a heterogeneous group in their skills [19] and internet skills represent a multidimensional concept [20]. Gaps in digital competence can be related to not only actual skills, but also a person's lack of confidence in his or her ability to use online services. Seo et al [21] showed that a lack of self-efficacy poses challenges for incarcerated people to access and use digital technologies. Consequently, this study focused on the significance of self-efficacy in the use of digital health care and social welfare services among incarcerated people in Finland. In particular, the study asked how incarcerated people's self-efficacy relates to their sense of control over the use of digital health care and social welfare services.

Use of Digital Health Care and Social Welfare Services by Incarcerated People in Finland

Incarcerated people in Finland are recognized as having a variety of social and health problems, and their ability to work is associated with major shortcomings [22]. The socioeconomic situation of people released from prison on parole is also quite weak, and according to the Criminal Sanctions Agency [23], about half (51%) are unemployed and only 26% are engaged in work. Furthermore, while 77% have a permanent home, 9% are homeless. The labor market position of people with a history of crime is weak in Finland, and is worse than in other Nordic countries [24]. The majority of incarcerated people in Finland have substance abuse problems [22]. Furthermore, according to Tuominen [25], neurocognitive and academic deficits are frequent among male incarcerated people, and the prevalence of these disorders is higher than seen in the general population. On the other hand, a Finnish prison survey (VASORA survey, 2018-2019, n=443) has shown that incarcerated people are a heterogeneous group from the perspective of perceived health status [26]. The key predictors of perceived health seem to be life expectations, a place to work or study, a permanent home, an adequate income, a permanent relationship, and a support person. In addition, the perceived health of incarcerated people in relation to other people of the same age deteriorates significantly with age. Overall, the need for health care and social welfare services for incarcerated people in Finland can be considered as significant, and in order to address these issues, incarcerated people should have access to digital social and health services already during their imprisonment and especially during their release phase.

In recent years, the digitalization of social and health care services has been rapid, both in Finland and in other Western countries. In practice, this has meant a rise in digital bookings, digital forms, virtual appointments, various forms of online therapy, and independently accessible databases and self-care programs. An important reform for the digitalization of Finnish social and health services has been the introduction of the *Kanta* service. The service includes *My Kanta* pages, where citizens can browse their own medical records and prescriptions, and reorder prescriptions online [27]. Especially with the COVID-19 pandemic, public mental health services and substance abuse treatment have also started to be provided virtually in Finland. In Finland, as part of the digital Health Village developed by university hospitals, *MentalHub* provides reliable information on mental health issues and services, tests, advice for self-care, and a variety of guidelines [28]. In addition to applying for electronic benefits, the various self-help programs and tests that are available and support substance abuse and mental health rehabilitation can be considered as important resources for incarcerated people.

In Finland, incarcerated people's access to the internet is regulated by the Imprisonment Act. According to the Act (Chapter 12, Section 9a [29]), prisoners may be authorized to use the internet for reasons of subsistence, work, education, justice, social affairs, and housing, or for other similarly important reasons. In a closed prison, the granting of a permit requires that the prisoner's access to nonlicensed websites is properly restricted. In the context of a prison, it is also required that any internet use does not pose a risk to the order or security of the prison, or to the safety of the prisoner or other persons.

In Finland, efforts have been made to facilitate incarcerated people's access to the internet and digital services through the Finnish Criminal Sanctions Agency's Smart Prison project, which follows the perspectives and goals that have emerged from discussions related to international smart corrections. The project aim is to develop the digitalization of activities and rehabilitation, and to adopt smart prison solutions in all prisons [30,31]. Currently, the development of digital services for prisons has focused specifically on the Hämeenlinna women's smart prison, which opened in autumn 2020, where functions and facilities have been designed with technology, in order to

XSL•FO
RenderX

support incarcerated people's integration into society and create a learning environment for a crime-free life. All incarcerated women have their own terminal in their prison cell, through which they can apply for a job or housing, participate in rehabilitation or remote education, and keep in touch with relatives or other officials through white-listed sites. The cell terminals were introduced in the spring of 2021, just after the data were collected for this study.

Self-efficacy as a Factor in the Use of Online Services

Research shows that incarcerated people tend to have deficient digital skills. This study addresses the digital self-efficacy of incarcerated people rather than their individual skills. Such a perspective considers that in addition to concrete skills, the incarcerated person's own experience of his or her ability is an important factor. According to Bandura [32], self-efficacy is the belief "in one's capabilities to organize and execute the courses of action required to produce given attainments." Self-efficacy is a key factor in human agency, and a considerable amount of research has been carried out on the effects of self-efficacy in areas such as motivation, health behavior, and learning.

The importance of self-efficacy has also been extensively studied among both incarcerated people and released people. Liem and Garcin [33] argued that self-efficacy appears to be a key element in postrelease success. Correspondingly, Forste et al [34] found that specific plans and perceptions of control and self-efficacy were all associated with intentions to stay out of trouble. According to Roth et al [35], academic self-efficacy is an important predictor for participation in prison education. There is also a discussion on criminal self-efficacy that refers to a person's perception of how capable they have been in successfully committing crimes [36]. Bailey and Ngwenyama [37] in turn emphasized the importance of self-efficacy when implementing community-based technological interventions among marginalized groups. The study by Allred et al [38] on courses that included college students (from outside the prison) and incarcerated people learning together in a prison classroom revealed the multidimensional nature of agency. Notably, they suggested that in addition to personal agency, we should look at the contextual elements of human agency. This highlights the importance of social support and cooperation in agency, but a lack of access to support can further aggravate the digital divide and increase digital exclusion [39].

A distinction can be made between general self-efficacy and specific self-efficacy such as computer or internet self-efficacy [40,41]. Internet self-efficacy has been shown to be a major factor in internet use [40], and perceived self-efficacy in computer use has been shown to have a significant positive influence on control beliefs [42]. However, the concept is not entirely unambiguous. According to Hsu and Chiu [41], general internet self-efficacy refers to "an individual's judgment of efficacy across multiple internet application domains," whereas web-specific self-efficacy relates to the use of a specific internet application or service. Their study shows that general internet self-efficacy influences the intention to use a digital service in the future through web-specific internet self-efficacy. Based on this previous research, it is important to distinguish between different forms of self-efficacy when looking at the use of digital health care and social welfare services among incarcerated people. In addition to general self-efficacy and internet self-efficacy, we need to analyze the specific sense of control associated with the use of digital health care and social welfare services. In accordance with the theory of planned behavior by Ajzen [43], we used the concept of perceived behavioral control, which, in this context, means the perception of the availability of skills, resources, and opportunities needed for using the technology concerned [44].

Previous studies have also found links between demographic factors, such as gender, age, and education level, and internet skills [20,45]. However, according to Noujaim et al [46], demographics and incarceration-related factors have very little effect on the health-related self-efficacy of incarcerated people. Likewise, Loeb et al [47] found that education and years of incarceration are not significantly related to self-efficacy. According to Chu [48], family support influences internet self-efficacy among older adults, and thus, it is reasonable to assume that it is also a significant factor in self-efficacy. Based on the studies described above, it is justified to analyze not only the effects of general self-efficacy and internet self-efficacy in the sense of control associated with the use of digital services, but also the effects of background factors such as age, education level, marital status, and number of convictions.

Aim and Hypotheses

This study examined self-efficacy and the sense of control over the use of digital health care and social welfare services among incarcerated people in Finland. The study asked the following question: How does incarcerated people's self-efficacy relate to their sense of control over the use of digital health care and social welfare services? In addition, the study investigated who incarcerated people receive help from when they encounter problems using digital health care and social welfare services.

Based on previous studies, we formulated the following 2 hypotheses: (1) Incarcerated people's sense of control over the use of digital health care and social welfare services depends on their general self-efficacy and internet self-efficacy and (2) Incarcerated people's internet self-efficacy depends on their general self-efficacy.

Methods

Sample

A cross-sectional study was conducted using a questionnaire. The study sought to attain regional coverage in Finland, and involved 11 of Finland's 26 prisons. The sample included 6 closed prisons and 5 open prisons. One of the closed prisons also had an open ward. In order to reach female prisoners, a closed women's prison and an open women's ward were included in the sample. A total of 225 people responded to the survey, and the response rate was 19.9% (225/1131).

Measures

General and Internet Self-efficacy

General self-efficacy was studied using the scale developed by Schwarzer and Jerusalem [49] that included 10 items. On the

XSL•FO
RenderX

scale, the possible responses to the questions are as follows: (1) not at all true, (2) hardly true, (3) moderately true, and (4) exactly true. Internet self-efficacy was studied using 8 Likert-type response questions drawn from the study by Eastin and LaRose [40]. The questions concern the use of internet hardware and software on a general level, and thus, they act as a specific measure of general internet self-efficacy. In this study, the response options for these items ranged from 1 (totally disagree) to 5 (totally agree).

Perceived Control

Five questions were prepared according to the concept of control by Ajzen [43]. These questions [50] were connected to the ease of use of digital services in general, perceived control of the use of digital health care and social welfare services, and confidence in learning to use new digital services. These questions were formed as statements and were evaluated on a Likert-type scale ranging from 1 (totally disagree) to 5 (totally agree).

Help With Digital Services

Respondents were also asked who they had received help from when they had problems using digital health care and social welfare services. In particular, the study examined how many of the respondents received help from friends, family members, prison employees, health care and social welfare employees, nongovernmental organizations (NGOs), religious organizations, and volunteers.

Sociodemographic Characteristics

The respondent's age was examined, as we can assume that a person's age has a negative association with digital technology use. Education level was measured using the following answer options: no primary school, primary school, secondary education, and higher education. The respondents were asked about their gender and marital status. In addition, their number of convictions, possession of bank IDs, and use of various health care and social welfare services during the past year were asked as background questions.

Procedure

The questionnaire used in the study was pretested using former incarcerated people (the training of experts by experience) as respondents (n=11). They were also asked questions related to the structure of the questionnaire, the ease of answering questions, the comprehensibility of the questions, and the clarity of the presented answer options. Based on the testing, minor changes were made to the layout of the form, but no changes were made to the questions themselves.

The data collection took place using a paper questionnaire between November 2020 and January 2021. The practice of data collection was agreed upon separately with each prison director. Data collection was carried out by a project worker (2 wards), university students working on the project (3 prisons), and prison staff. Responses were returned anonymously using sealed envelopes.

Statistical Analysis

The data were analyzed using IBM SPSS Statistics for Windows (version 27; IBM Corp). Averaged composite variables of individual questions were created, and their internal consistency was estimated using Cronbach α. The normality of distributions was checked using histograms, and they showed the distributions of composite variables as being close to normal.

The actual statistical analyses were conducted using the Pearson product-moment correlation coefficient, 2-tailed *t* test, and linear regression analysis with the Enter method. Before carrying out regression analyses, the validity of the conditions was checked. The normality of the residual distributions and the linearity condition were checked graphically, and the multicollinearity between the independent variables was examined using variance inflation factor coefficients. For regression analysis, dummy variables of marital status (1=married or common-law marriage; 0=unmarried or divorced) and education level (1=secondary or higher education; 0=no secondary education) were generated.

The indirect effects of general self-efficacy and the interaction between general self-efficacy and internet self-efficacy were analyzed using the Hayes bootstrapping method. Mediation and moderation analysis was carried out using PROCESS macro 3.5.3 for SPSS, and 5000 bootstrap samples were used. The indirect effect was considered significant when the upper and lower bounds (95% CIs) did not contain a value of zero.

Ethical Considerations

The study was conducted according to the ethical principles of the Finnish National Board on Research Integrity [51] and the WMA (World Medical Association) Declaration of Helsinki [52]. Ethical approval for the study was received from The Human Sciences Ethics Committee of the Helsinki Region Universities of Applied Sciences (decision 6/2020; September 25, 2020). Research approval was granted by the Criminal Sanction Agency (decision 30/332/2020; October 28, 2020). Informed consent was obtained from all incarcerated people who were involved in the study.

Results

Respondents

The average age of the respondents was 37.8 years (SD 11.7 years), and 8.9% (20/225) of the respondents were female. The representativeness of the survey turned out to be quite good. However, in some respects, the material was not fully representative. In particular, the level of education of the respondents was quite high, and the material included an overrepresentation of incarcerated people over the age of 50 years, as well as multiple recidivists with 10 or more convictions. Correspondingly, incarcerated people aged 30-39 years and those incarcerated for the first time were underrepresented. Open prisons were also overrepresented in the data (Table 1).

XSL•FO
RenderX

Table 1. Background data for respondents and the overall Finnish prison system.

Characteristic	Respondents (N=225), n (%)	Overall Finnish prison system, %
Sociodemographic factors		
Gender[a]		
Male	205 (91.1)	91.6
Female	20 (8.9)	8.4
Age[b]		
Under 30 years	58 (30.0)	29.7
30-39 years	53 (27.5)	35.0
40-49 years	45 (23.3)	20.7
50 years and over	37 (19.2)	14.6
Education level[c]		
No basic education	6 (3.1)	5.5
Basic education	78 (40.6)	55.5
Secondary education	86 (44.8)	35.4
Higher education	22 (11.5)	3.6
Marital status[d]		
Married	27 (12.7)	16.9
Common-law marriage	43 (20.3)	22.9
Divorced	54 (25.5)	19.9
Unmarried	88 (41.5)	37.0
Factors related to convictions		
Number of convictions[b]		
1	59 (33.3)	42.0
2	29 (16.4)	13.6
3-4	32 (18.1)	14.7
5-9	30 (16.9)	19.0
10 or over	26 (14.7)	10.7
Prison[a]		
Open (prison or ward)	117 (52.0)	35.6
Closed	108 (48.0)	64.4
Criminal sanctions region[a]		
Eastern and northern Finland	82 (36.4)	31.6
Southern Finland	83 (36.9)	35.4
Western Finland	60 (26.7)	33.0

[a]Reference value: All Finnish prisoners [23].

[b]Reference value: All Finnish sentenced prisoners [23].

[c]Reference value: Finnish parolees [23].

[d]Reference value: Data from the study on health, working capacity, and need for treatment of criminal sanction clients [22].

Use of Health Care and Social Welfare Services and Received Support

A total of 74.8% (166/222) of respondents answered that they use digital applications at least a few times a year. A large proportion (128/222, 57.7%) of respondents also said they searched for health information online, while a smaller proportion reported using self-care programs (87/225, 38.7%) or self-assessment tests (74/221, 33.5%) at least a few times a

XSL•FO
RenderX

year. Among the respondents, 10.8% (24/223) said that they do not use digital services.

Respondents said they received support for the use of digital health care and social welfare services, especially from friends (117/223, 52.5%) and family members (99/223, 44.4%). Among the respondents, 23.3% (52/223) had received help from prison employees, but only 13.5% (30/223) had received help from health care and social welfare employees and 5.4% (12/223) had received help from NGOs, religious organizations, or volunteers.

Composite Variables

Three averaged composite variables were constructed, and their reliability was seen as very high (α>.9) (Table 2).

The analysis of correlations (Table 3) showed that a person's perceived control over the use of digital health care and social welfare services, self-efficacy, and internet self-efficacy correlated significantly with each other. In particular, the dependence between control and internet self-efficacy was strong (r=0.708; P<.001). Age was negatively correlated with all of these 3 sum variables. In particular, a higher age seemed to be associated with lower internet self-efficacy (r=−0.421; P<.001) and poor perceived control over the use of digital health care and social welfare services (r=−0.261; P<.001). The number of convictions correlated negatively with internet self-efficacy (r=−0.293; P<.001) and perceived control (r=−0.182; P=.02), and as expected, the number of convictions also increased with age (r=0.414; P<.001).

Table 2. Data for composite variables.

Variable	Items, n	Respondents, n	Minimum value	Maximum value	Mean (SD)	Cronbach α
Self-efficacy	10	215	1.00	4.00	3.06 (0.61)	.932
Internet self-efficacy	8	218	1.00	5.00	3.77 (1.06)	.957
Perceived control	5	221	1.00	5.00	3.69 (1.04)	.908

Table 3. Pearson correlation analysis among the study variables.

Variable	Self-efficacy	Internet self-efficacy	Perceived control	Number of convictions	Age
Self-efficacy					
r	1	0.346	0.360	−0.023	−0.147
P value	—[a]	<.001	<.001	.77	.045
n	215	210	213	173	186
Internet self-efficacy					
r	0.346	1	0.708	−0.293	−0.421
P value	<.001	—	<.001	<.001	<.001
n	210	218	214	173	188
Perceived control					
r	0.360	0.708	1	−0.182	−0.261
P value	<.001	<.001	—	.02	<.001
n	213	214	221	175	191
Number of convictions					
r	−0.023	−0.293	−0.182	1	0.414
P value	.77	<.001	.02	—	<.001
n	173	173	175	177	161
Age					
r	−0.147	−0.421	−0.261	0.414	1
P value	.045	<.001	<.001	<.001	—
n	186	188	191	161	193

[a]Not applicable.

Incarcerated people serving their sentences in a closed prison or an open prison did not differ in terms of their self-efficacy (t_{213}=0.155; P=.88), internet self-efficacy (t_{216}=1.374; P=.17), or sense of control over digital health care and social welfare services (t_{219}=0.634; P=.53). In the data, no significant dependence was found between education level (secondary education or no secondary education) and self-efficacy (t_{181}=0.579; P=.56), internet self-efficacy (t_{185}=−0.012; P=.99),

XSL•FO
RenderX

or control over the use of digital health care and social welfare services (t_{187}=1.527; *P*=.13).

Self-efficacy and Internet Self-efficacy as Predictors of Perceived Control

Respondents' perceived control over the use of digital health care and social welfare services was examined using 5 questions. A total of 63.2% (141/223) of the respondents considered it easy to use digital services. Just over half (117/222, 52.7%) of the respondents believed that the use of various electronic social and health care services was completely under their control. The majority of respondents (155/222, 69.8%) were confident in their ability to easily learn how to use new digital services. Among the respondents, 71.7% (160/223) totally or partially agreed with the statement "I am able to apply for various social benefits electronically (eg, unemployment assistance, labor market assistance, sickness assistance, subsistence assistance, and housing assistance)." In contrast, less than half (104/223, 46.6%) totally or partially agreed with the statement "I am able to use various digital self-care programs."

A linear regression analysis was employed to examine the factors that explain perceived control over the use of digital health care and social welfare services. According to hypothesis 1, incarcerated people's sense of control over the use of digital health care and social welfare services depends on their general self-efficacy and internet self-efficacy. Thus, self-efficacy and internet self-efficacy were considered as independent variables. In addition, relevant background variables, such as age, level of education, marital status, and number of convictions, were included in the regression models. The first regression model (shown in Table 4) included all of the variables that were examined, and in the second model, the nonaffected variables (*P*>.05) were removed one at a time. Both models explained perceived control well (model 1: R^2=0.465; model 2: R^2=0.519). The analysis showed that internet self-efficacy offered the strongest explanation of perceived control, which was also consistent with the examination of correlates. In addition, according to model 2, general self-efficacy seemed to be positively associated with perceived control. Moderation analysis using the bootstrap method showed that the effect of the interaction term (general self-efficacy × internet self-efficacy) on the sense of control was not significant (coefficient=−0.032, 95% CI −0.156 to 0.092; *P*=.61).

Next, we examined what factors explained internet self-efficacy. In particular, we considered hypothesis 2, which states the following: Incarcerated people's internet self-efficacy depends on their general self-efficacy. In the regression analysis, self-efficacy was considered as an independent variable, and relevant background variables (age, level of education, marital status, and number of convictions) were also included in the first model (Table 5). In the second model, the nonaffected variables (*P*>.05) were removed one at a time. According to the analysis, general self-efficacy was associated with internet self-efficacy. In addition, age and number of convictions reduced internet self-efficacy. In contrast, level of education or marital status did not appear to affect internet self-efficacy or perceived control.

Mediation analysis revealed that general self-efficacy had both a direct effect and an indirect effect on perceived control over the use of digital health care and social welfare services, and internet self-efficacy was the mediating variable. According to bootstrapping analysis, the indirect effect coefficient was 0.391 (95% CI 0.202-0.577). General self-efficacy also had a significant direct effect on perceived control (indirect effect coefficient=0.253, 95% CI 0.076-0.430), but the indirect effect was stronger.

XSL•FO
RenderX

Table 4. Linear regression analysis with perceived control as the dependent variable.

Model	B[a]	SE[b]	β[c]	*t* (*df*)	*P* value
Model 1[d]					
Constant	0.920	0.529	N/A[e]	1.741 (128)	.08
Self-efficacy	0.179	0.119	0.101	1.495 (128)	.14
Internet self-efficacy	0.611	0.071	0.634	8.637 (128)	<.001
Age	0.003	0.007	0.034	0.419 (128)	.68
Number of convictions	−0.010	0.015	−0.045	−0.625 (128)	.53
Education level[f]	−0.239	0.139	−0.120	−1.723 (128)	.09
Marital status[f]	−0.075	0.146	−0.035	−0.516 (128)	.61
Model 2[g]					
Constant	0.396	0.284	N/A	1.396 (205)	.16
Self-efficacy	0.253	0.090	0.145	2.814 (205)	.005
Internet self-efficacy	0.661	0.052	0.658	12.765 (205)	<.001

[a]Unstandardized regression coefficient.

[b]SE: standard error.

[c]Standardized regression coefficient.

[d]R^2=0.465; adjusted R^2=0.440; $F_{6,128}$=18.5, N=135; P<.001.

[e]N/A: not applicable.

[f]Dummy.

[g]R^2=0.519; adjusted R^2=0.514; $F_{2,205}$=110.6, N=208; P<.001.

Table 5. Linear regression analysis with internet self-efficacy as the dependent variable.

Model	B[a]	SE[b]	β[c]	*t* (*df*)	*P* value
Model 1[d]					
Constant	3.916	0.558	N/A[e]	7.019 (130)	<.001
Self-efficacy	0.361	0.145	0.196	2.493 (130)	.01
Age	−0.033	0.008	−0.364	−4.097 (130)	<.001
Number of convictions	−0.032	0.019	−0.144	−1.728 (130)	.09
Education level[f]	0.123	0.169	0.059	0.729 (130)	.47
Marital status[f]	0.091	0.180	0.040	0.504 (130)	.62
Model 2[g]					
Constant	3.408	0.495	N/A	6.879 (152)	<.001
Self-efficacy	0.490	0.125	0.279	3.938 (152)	<.001
Age	−0.027	0.007	−0.295	3.851 (152)	<.001
Number of convictions	−0.035	0.017	−0.159	2.100 (152)	.04

[a]Unstandardized regression coefficient.

[b]SE: standard error.

[c]Standardized regression coefficient.

[d]R^2=0.232; adjusted R^2=0.203; $F_{5,130}$=7.86, N=136; P<.001.

[e]N/A: not applicable.

[f]Dummy.

[g]R^2=0.251; adjusted R^2=0.237; $F_{3,152}$=17.0, N=156; P<.001.

XSL•FO
RenderX

Discussion

Principal Findings

Digital health care and social welfare services play important roles in the planned release from prison and integration into society. However, in prisons, the use of the internet and digital services is limited, and many psychological factors create barriers for their use. This study highlighted the importance of self-efficacy from the perspective of digital service use. In addition to the specific self-efficacy associated with the use of the internet, general self-efficacy also has a significant effect on the sense of control over using digital health care and social welfare services.

The first hypothesis that "incarcerated people's sense of control over the use of digital health care and social welfare services depends on their general self-efficacy and internet self-efficacy" is supported. The second hypothesis that "incarcerated people's internet self-efficacy depends on their general self-efficacy" is also supported. The study further found that older incarcerated people and multiple recidivists see their skills in using digital services as poor.

Reflection on the Results

In terms of the use of digital systems, the sense of control associated with such use is central [44], and depends strongly in turn on internet self-efficacy. Thus, when supporting the use of digital health care and social welfare services, general information technology skills should be taken into account in order to strengthen the individual's internet self-efficacy. Particularly, it is not enough to only support incarcerated people in their use of individual services, but it is also essential that people learn to trust their ability to operate online. Furthermore, Reisdorf and Rikard [7] pointed out that appropriate digital skills training during re-entry could mitigate some of the adverse effects of digital exclusion during incarceration and thus facilitate the re-entry of incarcerated people.

While this study did not directly focus on the process of desistance of a person released from prison, it highlighted the importance of digital services for the reintegration of released people and their perceived control over the use of digital services during their imprisonment. Previous studies have shown that incarcerated people should have access to digital services during their imprisonment, in order to facilitate their integration into society [7,9]. In particular, research has shown that both offline and online rehabilitation measures should be implemented during imprisonment and in the re-entry phase [7]. The digitalization of services and facilities makes it possible for incarcerated people to use digital health care and social welfare services from prison, to participate in distance learning and rehabilitation, and to communicate with family or different authorities, and it facilitates their search for employment [13,14].

When supporting incarcerated people in the use of digital services, the importance of general self-efficacy should also be identified. Persons with poor general life management skills often lack the skills to use digital health care and social welfare services. For this marginalized group of people, it is not enough to pay attention to only the use of the internet or digital health care and social welfare services, and one needs to pay attention to life management skills in general. This means not only facilitating support for substance abuse and mental health problems, but also introducing a range of support measures for use in everyday life [19].

Seo et al [21] pointed out that a lack of self-efficacy poses challenges for incarcerated people in accessing and using digital technologies. Our results are consistent with this, and the results show that without adequate self-efficacy and internet self-efficacy, released people are at risk of digital exclusion. Digital exclusion and social exclusion are furthermore mutually reinforcing processes that can, however, be moderated by appropriate digital skills, technology-positive attitudes, and good access to digital services [15,16].

According to previous studies, incarceration-related factors [46] or the number of years of incarceration [47] have no significant effect on health-related self-efficacy. Consistent with this, this study found no link between the number of convictions and self-efficacy, and in contrast, the number of convictions was negatively correlated with internet self-efficacy. The perceived digital ability of multiple recidivists appeared to be lower than that of other incarcerated people. This is particularly challenging because they also have a high risk of recidivism [53]. Therefore, in the prevention of recidivism, special attention should be paid to strengthening the digital skills and internet self-efficacy of recidivists.

Previous studies have also found links between demographic factors and digital skills [20,45], but the findings have been contradictory [46,47]. The importance of sociodemographic factors, such as education level and marital status, appeared to be surprisingly low in the data. Overall, a feeling of self-efficacy was more essential than an individual's sociodemographic factors. However, older prisoners had a low level of internet self-efficacy and thus needed special support for accessing the internet and digital services.

In general, respondents estimated that they have good control over their use of digital health care and social welfare services. In particular, making electronic applications was considered to be easy. However, respondents felt that their skills relating to self-care programs were weak, even though there were many opportunities associated with self-care programs. For example, substance abuse treatment services in prisons currently reach only some of those who need them, and digital self-care programs can significantly increase treatment coverage. Accordingly, special attention should be paid to the ease of use of self-care programs and teaching their use in the future.

Previous studies have emphasized the importance of support from employees [19] and close ones [39,45,48] for digital use by incarcerated people. According to this study, the importance of informal support from family and friends for the use of digital health care and social welfare services was significant. Several respondents reported receiving support from prison employees for their use of digital services, but there is still much room for improvement. Unexpectedly, the role of NGOs as providers of digital support seemed to be quite small among the respondents of this study.

XSL•FO
RenderX

Limitations

The data for this study were collected from different prisons across Finland, and the representativeness of the data can be considered to be quite good. In addition, the reliability of all studied variables was high. However, there were some limitations related to the research design and data. First, since this was a cross-sectional study, the analysis showed the statistical relationships between the variables, but not the causal relationships. Second, less than 10% of incarcerated people in Finland participated in the study, and thus, the results cannot be directly generalized. As a third consideration, there were some differences in background variables between the respondents in this study and the overall prison population in Finland. Specifically, the material overrepresented people serving their sentences in an open prison, and the level of education of the respondents was also quite high. On the other hand, based on our analysis, the effect of these factors appears to be reasonably small in terms of perceived control and internet self-efficacy. However, the question of the challenges of using digital services faced by the least educated incarcerated people (ie, those with no basic education) was excluded from this research design.

As a fourth area of consideration, the findings relate to the perceived control over digital health care and social welfare services, but not to actual digital skills or the actual use of digital services. Furthermore, the study examined the self-efficacy and internet self-efficacy of incarcerated people through self-assessments, and as respondents were asked to answer questions from the perspective of their current situation (ie, the situation during their time in prison), the study does not reveal how self-efficacy or internet self-efficacy changes after release.

Conclusions

The digitalization of prisons offers many opportunities in terms of improving access to health care and social welfare services, and increasing the efficiency of prison processes. On the other hand, the technology chosen to bring about such improvements should also address the diverse challenges of the prison environment and recognize the various risk factors that present in this unique environment. Special consideration should also be given to how digitization and development work responds to the needs of incarcerated people for their integration into society and the prevention of recidivism. In order for incarcerated people to have an opportunity to adapt to life in a digital society, access to the internet and to digital health care and social welfare services must be made available during their imprisonment. In addition, during imprisonment, attention should be paid to strengthening incarcerated people's digital skills, with support provided by prison employees. Older prisoners and multiple returnees, in particular, tend to be excluded from the digital society. The results of this study recommend paying attention to not only developing digital skills, but also strengthening the general self-efficacy of incarcerated people. Inherent to this is the key role played by social support, training, and learning of various skills needed to pursue an everyday crime-free life.

Acknowledgments

This research was conducted as part of the "Toward a Socially Inclusive Digital Society: Transforming Service Culture" project (DigiIN) and was funded by the Strategic Research Council at the Academy of Finland (grant numbers 327169 and 327145).

Conflicts of Interest

None declared.

References

1. Farrall S, Bottoms A, Shapland J. Social structures and desistance from crime. European Journal of Criminology 2010 Oct 21;7(6):546-570. [doi: 10.1177/1477370810376574]
2. Kazemian L. Desistance From Crime. Journal of Contemporary Criminal Justice 2016 Jul 24;23(1):5-27. [doi: 10.1177/1043986206298940]
3. Laub JH, Sampson RJ. Understanding Desistance from Crime. Crime and Justice 2001 Jan;28:1-69. [doi: 10.1086/652208]
4. Shapland J, Bottoms A, Farrall S, McNeill F, Priede C, Robinson G. The quality of probation supervision - a literature review. University of Sheffield. URL: https://www.sheffield.ac.uk/polopoly_fs/1.159010!/file/QualityofProbationSupervision.pdf [accessed 2022-01-24]
5. Farrall S, Calverley A. Understanding desistance from crime: Theoretical directions in resettlement and rehabilitation. Open University Press. URL: http://www.8-926-145-87-01.ru/wp-content/uploads/2014/12/Stephen-Farrall-Adam-Calverley-Understanding-desistance-from-crime-Crime-and-Justice-20051.pdf [accessed 2022-01-24]
6. McCulloch T. Probation, social context and desistance: Retracing the relationship. Probation Journal 2016 Jun 25;52(1):8-22. [doi: 10.1177/0264550505050623]
7. Reisdorf BC, Rikard RV. Digital Rehabilitation: A Model of Reentry Into the Digital Age. American Behavioral Scientist 2018 May 04;62(9):1273-1290. [doi: 10.1177/0002764218773817]
8. McNeill F. Supporting Desistance in Probation Practice: A Response to Maruna, Porter and Carvalho. Probation Journal 2016 Jun 25;51(3):241-247. [doi: 10.1177/0264550504046746]
9. Reisdorf BC, Jewkes Y. (B)Locked sites: cases of Internet use in three British prisons. Information, Communication & Society 2016 Mar 02;19(6):771-786. [doi: 10.1080/1369118x.2016.1153124]

XSL•FO
RenderX

10. Knight V, Van De Steene S. Digitizing the prison: The light and dark future. Prison Service Journal 2017;231:22-30 [FREE Full text]
11. Knight V, van de Steene S. The Capacity and Capability of Digital Innovation in Prisons: Towards Smart Prisons. Advancing Corrections 2017:88-101 [FREE Full text]
12. Robberechts J, Beyens K. PrisonCloud: The beating heart of the digital prison cell. In: Turner J, Knight V, editors. The Prison Cell. Cham: Palgrave Macmillan; 2020:283-303.
13. Knight V. Some Observations on the Digital Landscape of Prisons Today. Prison Service Journal 2015;220:3-10 [FREE Full text]
14. Ogbonnaya-Ogburu I, Toyama K, Dillahunt T. Towards an Effective Digital Literacy Intervention to Assist Returning Citizens with Job Search. In: CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: 2019 CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland, UK. [doi: 10.1145/3290605.3300315]
15. Helsper EJ. A Corresponding Fields Model for the Links Between Social and Digital Exclusion. Commun Theor 2012 Oct 15;22(4):403-426. [doi: 10.1111/j.1468-2885.2012.01416.x]
16. Helsper E. Digital inclusion: an analysis of social disadvantage and the information society. LSE Research Online. London, UK: Department for Communities and Local Government; 2008. URL: http://eprints.lse.ac.uk/26938/ [accessed 2022-01-24]
17. Monteiro A, Barros R, Leite C. Lifelong learning through e-learning in European prisons: Rethinking digital and social inclusion. IATED. 2015. URL: https://library.iated.org/view/MONTEIRO2015LIF [accessed 2022-01-24]
18. Jewkes Y, Reisdorf BC. A brave new world: The problems and opportunities presented by new media technologies in prisons. Criminology & Criminal Justice 2016 Jul 24;16(5):534-551. [doi: 10.1177/1748895816654953]
19. Järveläinen E, Rantanen T. Incarcerated people's challenges for digital inclusion in Finnish prisons. Nordic Journal of Criminology 2020 Sep 14:1-20. [doi: 10.1080/2578983x.2020.1819092]
20. van Deursen AJAM, Courtois C, van Dijk JAGM. Internet Skills, Sources of Support, and Benefiting From Internet Use. International Journal of Human-Computer Interaction 2014 Mar 07;30(4):278-290. [doi: 10.1080/10447318.2013.858458]
21. Seo H, Britton H, Ramaswamy M, Altschwager D, Blomberg M, Aromona S, et al. Returning to the digital world: Digital technology use and privacy management of women transitioning from incarceration. New Media & Society 2020 Oct 20;24(3):641-666. [doi: 10.1177/1461444820966993]
22. Joukamaa M. Rikosseuraamus- asiakkaiden terveys, työkyky ja hoidontarve. Rikosseuraamus. URL: https://www.rikosseuraamus.fi/material/attachments/rise/julkaisut-risenjulkaisusarja/6AqMACEr8/RISE_1_2010_Rikosseuraamusasiakkaiden_terveys_tyokyky_ja_hoidontarve.pdf [accessed 2022-01-24]
23. Rikosseuraamuslaitoksen monisteita 2/2020: Rikosseuraamusasiakkaat 1.5.2020. Rikosseuraamus. URL: https://www.rikosseuraamus.fi/material/attachments/rise/julkaisut-monisteetjaraportit/YTPwEq9MH/Rikosseuraamusasiakkaat_1.5.2020_tehty_29.6.2020.pdf [accessed 2022-01-24]
24. Aaltonen M, Skardhamar T, Nilsson A, Højsgaard Andersen L, Bäckman O, Estrada F, et al. Comparing Employment Trajectories before and after First Imprisonment in Four Nordic Countries. CRIMIN 2016 Feb 28:azw026. [doi: 10.1093/bjc/azw026]
25. Tuominen T. Neurocognitive deficits, academic difficulties and substance dependence among Finnish offenders: connections to recidivism and implications for rehabilitation. Turun Yliopiston Julkaisuja – Annales Universitatis Turkuensis. 2018. URL: https://www.utupub.fi/bitstream/handle/10024/144773/AnnalesB453Tuominen.pdf?se [accessed 2022-01-24]
26. Kauhanen J. Vankien koettu terveys [Perceived health of prisoners]. In: Kainulainen S, Saari J, editors. Suomalainen vanki — Arjen rakenteet ja elämänlaatu vankilassa. Tampere, Finland: Vastapaino; 2021:57-75.
27. What are the Kanta Services? Kanta. URL: https://www.kanta.fi/en/system-developers/what-are-kanta-services [accessed 2022-01-24]
28. Health Village. MentalHub. URL: https://www.mielenterveystalo.fi/en/Pages/default.aspx [accessed 2022-01-24]
29. The Imprisonment Act (767/2005). Finlex. URL: https://www.finlex.fi/fi/laki/kaannokset/2005/en20050767.pdf [accessed 2022-01-24]
30. Smart Prison Project. Rikosseuraamus. URL: https://rikosseuraamus.fi/en/index/topical/projects.html [accessed 2022-01-24]
31. Puolakka P. Smart Prison. EuroPris. 2020. URL: https://www.europris.org/wp-content/uploads/2020/10/Pia-Puolakka-Smart-Prisons-Finland.pdf [accessed 2022-01-24]
32. Bandura A. Self-efficacy: The exercise of control. New York, NY: Freeman; 1997.
33. Liem M, Garcin J. Post-Release Success among Paroled Lifers. Laws 2014 Dec 15;3(4):798-823. [doi: 10.3390/laws3040798]
34. Forste R, Clarke L, Bahr S. Staying Out of Trouble: Intentions of Young Male Offenders. Int J Offender Ther Comp Criminol 2010 Jan 28;55(3):430-444. [doi: 10.1177/0306624x09359649]
35. Roth B, Westrheim K, Jones L, Manger T. Academic Self-Efficacy, Educational Motives and Aspects of the Prison Sentence as Predictors for Participation in Prison Education. The Journal of Correctional Education 2017;68(3):19-40 [FREE Full text]
36. Brezina T, Topalli V. Criminal Self-Efficacy. Criminal Justice and Behavior 2012 Apr 09;39(8):1042-1062. [doi: 10.1177/0093854812438345]

XSL•FO
RenderX

37. Bailey A, Ngwenyama O. Community Mediation through ICTs: Seeking to Bridge Digital and Community Divides. Journal of Community Informatics 2016 Mar 27;12(1):69-89. [doi: 10.15353/joci.v12i1.3241]
38. Allred SL, Harrison LD, O'Connell DJ. Self-Efficacy. The Prison Journal 2013 Jan 24;93(2):211-233. [doi: 10.1177/0032885512472964]
39. Helsper EJ, van Deursen AJAM. Do the rich get digitally richer? Quantity and quality of support for digital engagement. Information, Communication & Society 2016 Jun 29;20(5):700-714. [doi: 10.1080/1369118x.2016.1203454]
40. Eastin M, LaRose R. Internet Self-Efficacy and the Psychology of the Digital Divide. Journal of Computer-Mediated Communication 2000;6(1):JCMC611. [doi: 10.1111/j.1083-6101.2000.tb00110.x]
41. Hsu M, Chiu C. Internet self-efficacy and electronic service acceptance. Decision Support Systems 2004 Dec;38(3):369-381. [doi: 10.1016/j.dss.2003.08.001]
42. Melzner J, Heinze J, Fritsch T. Mobile Health Applications in Workplace Health Promotion: An Integrated Conceptual Adoption Framework. Procedia Technology 2014;16:1374-1382. [doi: 10.1016/j.protcy.2014.10.155]
43. Ajzen I. Nature and operation of attitudes. Annu Rev Psychol 2001;52:27-58. [doi: 10.1146/annurev.psych.52.1.27] [Medline: 11148298]
44. Holden RJ, Karsh B. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
45. Courtois C, Verdegem P. With a little help from my friends: An analysis of the role of social support in digital inequalities. New Media & Society 2016 Jul 09;18(8):1508-1527. [doi: 10.1177/1461444814562162]
46. Noujaim D, Fortinsky RH, Barry LC. The Relationship Between Emotional Support and Health-Related Self-Efficacy in Older Prisoners. J Aging Health 2019 Mar 27;31(3):439-462. [doi: 10.1177/0898264317733363] [Medline: 29254413]
47. Loeb S, Steffensmeier D, Kassab C. Predictors of self-efficacy and self-rated health for older male inmates. J Adv Nurs 2011 Apr;67(4):811-820 [FREE Full text] [doi: 10.1111/j.1365-2648.2010.05542.x] [Medline: 21198807]
48. Chu RJ. How family support and Internet self-efficacy influence the effects of e-learning among higher aged adults – Analyses of gender and age differences. Computers & Education 2010 Aug;55(1):255-264. [doi: 10.1016/j.compedu.2010.01.011]
49. Schwarzer R, Jerusalem M. Generalized Self-Efficacy scale. In: Weinman J, Wright S, Johnston M, editors. Measures in health psychology: A user's portfolio. Causal and control beliefs. Windsor, UK: NFER-Nelson; 1995:35-37.
50. Rantanen T, Järveläinen E, Leppälahti T. Prisoners as Users of Digital Health Care and Social Welfare Services: A Finnish Attitude Survey. Int J Environ Res Public Health 2021 May 21;18(11):e68053 [FREE Full text] [doi: 10.3390/ijerph18115528] [Medline: 34064026]
51. Finnish National Board on Research Integrity TENK guidelines 2019: The ethical principles of research with human participants and ethical review in the human sciences in Finland. The Finnish National Board on Research Integrity TENK. URL: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf [accessed 2022-01-24]
52. WMA (World Medical Association) Declaration of Helsinki-Ethical Principles for Medical Research Involving Human Subjects. WMA. URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [accessed 2022-01-24]
53. Criminal Sanctions Agency Statistical Yearbook 2019. Rikosseuraamus. URL: https://www.rikosseuraamus.fi/material/attachments/rise/julkaisut-tilastollinenvuosikirja/pjeawUKaf/Statistical_Yearbook_2019_of_the_Criminal_Sanctions_Agency.pdf [accessed 2022-01-24]

Abbreviations

NGO: nongovernmental organization

Edited by A Mavragani; submitted 26.01.22; peer-reviewed by V Knight, C Li; comments to author 25.04.22; revised version received 02.05.22; accepted 02.05.22; published 31.05.22.

Please cite as:
Rantanen T, Järveläinen E, Leppälahti T
Self-efficacy and Use of Digital Health Care and Social Welfare Services Among Incarcerated People: Cross-sectional Survey Study
J Med Internet Res 2022;24(5):e36799
URL: https://www.jmir.org/2022/5/e36799
doi:10.2196/36799
PMID:35639446

©Teemu Rantanen, Eeva Järveläinen, Teppo Leppälahti. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 31.05.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution

XSL•FO
RenderX

License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Web-Based Independent Versus Laboratory-Based Stop-Signal Task Performance: Within-Subjects Counterbalanced Comparison Study

Antoinette Poulton[1], BA, MA, PhD; Li Peng Evelyn Chen[1], MPsych; Gezelle Dali[1], BSc; Michael Fox[1], BSc, MSc; Robert Hester[1], PhD

Melbourne School of Psychological Sciences, University of Melbourne, Parkville, Australia

Corresponding Author:
Antoinette Poulton, BA, MA, PhD
Melbourne School of Psychological Sciences
University of Melbourne
Redmond Barry Building
Parkville, 3010
Australia
Phone: 61 3 8344 6377
Fax: 61 3 9347 6618
Email: antoinette.poulton@unimelb.edu.au

Abstract

Background: Considered a facet of behavioral impulsivity, response inhibition facilitates adaptive and goal-directed behavior. It is often assessed using the Stop-Signal Task (SST), which is presented on stand-alone computers under controlled laboratory conditions. Sample size may consequently be a function of cost or time and sample diversity constrained to those willing or able to attend the laboratory. Statistical power and generalizability of results might, in turn, be impacted. Such limitations may potentially be overcome via the implementation of web-based testing.

Objective: The aim of this study was to investigate if there were differences between variables derived from a web-based SST when it was undertaken independently—that is, outside the laboratory, on any computer, and in the absence of researchers—versus when it was performed under laboratory conditions.

Methods: We programmed a web-based SST in HTML and JavaScript and employed a counterbalanced design. A total of 166 individuals (mean age 19.72, SD 1.85, range 18-36 years; 146/166, 88% female) were recruited. Of them, 79 undertook the independent task prior to visiting the laboratory and 78 completed the independent task following their laboratory visit. The average time between SST testing was 3.72 (SD 2.86) days. Dependent samples and Bayesian paired samples *t* tests were used to examine differences between laboratory-based and independent SST variables. Correlational analyses were conducted on stop-signal reaction times (SSRT).

Results: After exclusions, 123 participants (mean age 19.73, SD 1.97 years) completed the SST both in the laboratory and independently. While participants were less accurate on go trials and exhibited reduced inhibitory control when undertaking the independent—compared to the laboratory-based—SST, there was a positive association between the SSRT of each condition (r=.48; P<.001; 95% CI 0.33-0.61).

Conclusions: Findings suggest a web-based SST, which participants undertake on any computer, at any location, and in the absence of the researcher, is a suitable measure of response inhibition.

(J Med Internet Res 2022;24(5):e32922) doi:10.2196/32922

KEYWORDS

Stop-Signal Task; response inhibition; inhibitory control; online assessment; web-based assessment; cognition

XSL•FO
RenderX

Introduction

Considered a facet of behavioral impulsivity, response inhibition refers to the capacity to withhold, interrupt, or delay a prepotent behavioral response and is a key element of executive function [1-4]. Also termed "inhibitory control" or cognitive control," it facilitates adaptive and goal-directed behavior [5]. The Stop-Signal Task (SST), a commonly employed measure of response inhibition [6], has been used to examine the inhibitory control of healthy adults and children [7-9], older adults [10], and clinical groups [11]. It is routinely used to show how individuals diagnosed with attention-deficit/hyperactivity disorder (ADHD) [12] and substance use disorders (SUDs) [13,14] tend to be characterized by heightened impulsivity. While there is abundant literature examining the psychometrics of the task when it is undertaken on dedicated computers under laboratory conditions [15-17], it is unclear if performance on web-based versions of the task differs as a function of the testing environment.

Although there are several variants of the SST, they all fundamentally assess the ability to suppress a motor response that has already been initiated [18-21]. In all cases, individuals must respond rapidly to frequently appearing (go) stimuli but inhibit responses to others (stop signals) presented much less often [17,21-23]. The imbalance in the occurrence of each type of stimulus creates a response prepotency that manifests in a difficulty inhibiting responses when required. The go component of the SST is essentially a 2-choice reaction time (RT) task that involves the electronic presentation of 1 of 2 stimuli (for example, X or O). In response, participants are required to press the corresponding letter on a keyboard as quickly as possible—this generates a go RT. The stop component of the task typically occurs on 25% of trials and comprises the presentation of a stop signal—in the form of an auditory tone or visual indicator—designed to inform participants that they must withhold (or inhibit) their response to the stimulus on that trial. The period between the presentation of the go stimulus and the stop signal is known as the stop-signal delay (SSD) [21,24]. Although initially usually set at 250 milliseconds, the onset of the SSD varies dynamically in a stepwise manner on each trial and as a function of participant performance. In this way, successful inhibition approaches 50% accuracy by the end of the task. Stop-signal reaction time (SSRT) is often the main variable of interest in the SST and represents the difference between mean go RT and the average SSD [20,21].

The SST is typically programmed using common software packages and has traditionally been presented on stand-alone computers in controlled research settings [23]. This assists in ensuring that task presentation is consistent across participants and variability—related both to computer hardware or software and the testing environment—is minimal. Participants are thus generally required to visit the laboratory in order to take part in studies using this tool. They may even be tested individually. This gives rise to two potential limitations: sample size becomes a function of cost or time constraints, and sample diversity is restricted to those willing and able to attend the laboratory. In turn, this may impact power and means findings may not be generalizable to the wider population. Moreover, COVID-19 restrictions have meant that in-person testing is frequently unavailable or hampered by the need to implement social distancing, cleaning or sanitizing, and personal protective equipment protocols. While such procedures may have unintended consequences that impact the quality of the data, they are also likely to be costly and time-consuming [25]. This may further exacerbate sample size and diversity issues.

These limitations may potentially be overcome through web-based testing. While there has been a substantial increase in the popularity of using the internet as a medium for conducting research in social psychology—which appears to have resulted in larger samples and increased statistical power [26]—this trend has been less evident in the cognitive arena, possibly owing to validity and reliability concerns [27,28]. Nonetheless, researchers have, more recently, begun to examine whether participants perform in similar ways when undertaking web-based cognitive tasks independently versus in the laboratory [29-32]. Results of these studies suggest that while main effects remain the same, there might be some timing and accuracy offsets related to participant concentration and hardware or software variability in uncontrolled testing environments [29,31]. To date, the SST has not been the subject of such an investigation. The aim of this study was therefore to investigate whether performance on a web-based version of the SST differed as a function of the testing environment. Data were collected prior to the onset of the COVID-19 pandemic. Informed by findings in similar previous studies [29,31], we hypothesized the independent web-based SST—that is, the SST performed outside the laboratory, on any computer, and in the absence of any researcher—would be characterized by decreased go and stop accuracy, increased go omissions and go errors, and longer SSRTs, as compared to the laboratory-based SST carried out on the web. Additionally, relative to the laboratory-based SST, intraindividual variability would be greater in the independent task. Nonetheless, given that other studies have also found acceptable comparability between independent and laboratory-based cognitive tasks [29-32], we expected that there would be a robust positive relationship between independent and laboratory-based SSRTs.

Methods

Recruitment and Procedure

Participants were 166 individuals (mean age 19.72, SD 1.85, range 18-36 years; 146/166, 88% female) who completed this study as part of their undergraduate psychology studies. First-year psychology students at the University of Melbourne are encouraged to take part in studies being conducted within the School of Psychological Sciences. Students receive course credit as reimbursement for their time.

On signing up for the study (via a School of Psychological Sciences research participation landing page), participants were randomly assigned to either first complete the web-based SST in the laboratory or independently. The independent condition was thus completed in counterbalanced order, with half of the participants undertaking the task prior to visiting the laboratory and the other half doing it following their laboratory visit. In both cases, consent was obtained via a web-based form, and

XSL•FO
RenderX

links to the task were emailed to the participants. In the laboratory, participants also completed alcohol and substance use surveys.

Measures

Substance Use

Participants completed the Alcohol Use Disorders Identification Test (AUDIT), which assesses alcohol intake, problems, and dependence with reference to the preceding 6 months [33]. Harmful use of licit and illicit drugs was assessed using the Alcohol, Smoking and Substance Involvement Screen Test (ASSIST), which assesses frequency of use and associated problems over the previous 3 months [34].

Inhibitory Control

The web-based SST was programmed using HTML (version 5) and JavaScript client-side along with PHP and MySQL server-side for data storage and management [35]. The task is run on Windows or Mac desktop or laptop computers and supported by all major browsers. Initial instructions are provided across 2 screens in a white 20-point Sans Serif font on a black background (Figure 1). The task consists of a practice block of 32 trials and 3 blocks of 64 experimental trials. Practice trials have an intertrial interval (ITI) of 4250 milliseconds, while experimental trials have ITI of 2250 milliseconds (Figure 2). The fixation cross and stimuli are rendered in 100- and 150-point Serif font, respectively. Trial-by-trial feedback is provided during the practice block, while block-based feedback is given during experimental trials. Within-task prompts or feedback are provided in white and colored 25-point Serif font on a black background. During the practice block, the following trial-based prompts or feedback are provided:

- Fixation cross screen: *Get ready*
- Go stimuli presentation screen: *Press X (or O) as fast as possible*
- Stop stimuli presentation screen: *Do not press any key*
- Successful go with response time < 500 milliseconds: *Hit* (green text)
- Successful go with response time ≥ 500 milliseconds: *Hit (but try to go faster)* (yellow text)
- Go omission: Miss *(you must go faster)* (red text)
- Incorrect go: Miss *(incorrect keystroke)* (red text)
- Following three consecutive go omissions: *WARNING: You MUST respond to X/O go stimuli as fast as possible* (red text)
- Successful inhibition on a stop trial: *Successful stop – Well done!* (green text)
- Unsuccessful inhibition on a stop trial: *Unsuccessful stop – try not to respond to stop trials!* (red text)
- Blank screen: *Wait*

During experimental blocks, the only trial-based prompt provided occurs if participants neglect to respond to 3 consecutive go trials. In this case, participants are warned: *You MUST respond to X/O go stimuli as fast as possible* (red text). At the end of both the practice and experimental blocks, participants are provided with the following block-based feedback: number of incorrect responses to go stimuli; number of missed responses to go stimuli; mean reaction time to go stimuli (where this is ≥500 milliseconds, participants are warned *Too slow! Respond faster*); percentage of correctly suppressed responses on stop trials; and seconds left to wait (10-second countdown to the next block).

Go stimuli comprise random presentation of letters X or O that map to corresponding keyboard letters. A stop signal in the form of a white box surrounding the go stimuli appears on 25% of randomly selected trials. Stop signals are not presented on consecutive trials. The initial SSD is set at 250 milliseconds and adjusts dynamically as a function of the participant response; successful inhibitions result in a 50-millisecond increase in the SSD, while unsuccessful inhibitions decrease it by 50 milliseconds. Variables of interest may include go accuracy, omissions, and errors; average go RT; intraindividual SD; stop accuracy; mean SSD; average RT on unsuccessful stop trials; and SSRT [21]. SSRT is derived when the mean SSD is subtracted from average go RT; greater SSRTs denote reduced inhibition ability [22]. Participants are excluded if the mean RT of either correct or incorrect failed stops (ie, failed stops where the key press does or does not respectively accord with the stimulus) is greater than the mean go RT [21]. They are also excluded if the stop accuracy is less than 25% or greater than 75%, go errors are greater than 10%, or if SSRT is less than 50 milliseconds [17].

XSL•FO
RenderX

Figure 1. Initial instructions (provided across 2 screens) for the web-based Stop-Signal Task.

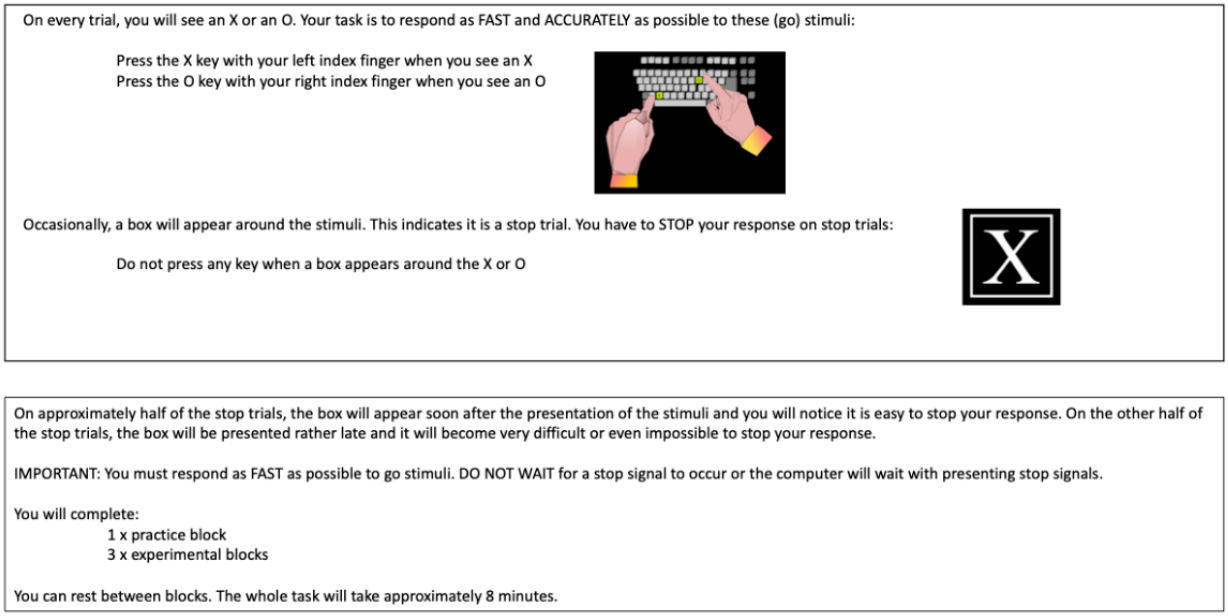
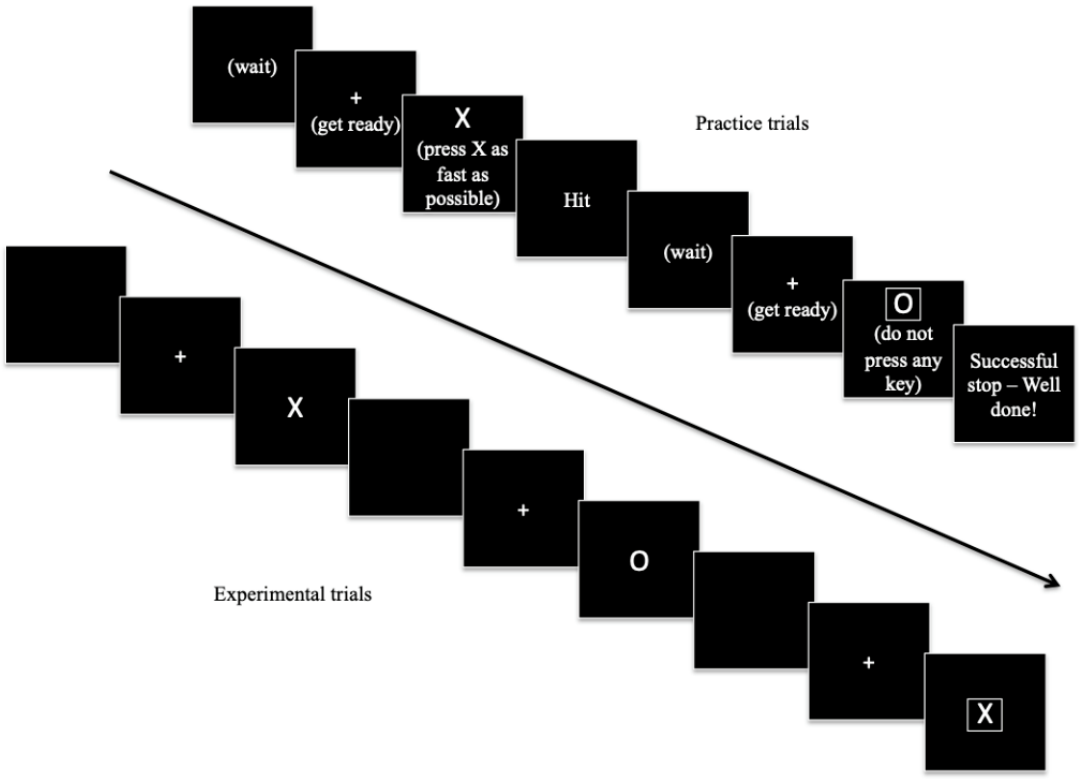


Figure 2. Schematic representation of Go and Stop trials in the practice and experimental blocks of the web-based Stop-Signal Task. Practice trials have an inter-trial interval (ITI) of 4250 milliseconds and comprise a blank screen (1000 milliseconds), fixation cross (250 milliseconds), stimulus presentation (1000 milliseconds), and feedback screen (2000 milliseconds). Experimental trials have an ITI of 2250 milliseconds and comprise a blank screen (1000 milliseconds), fixation cross (250 milliseconds), and stimulus presentation (1000 milliseconds).



Data Analysis

To achieve a medium to small effect with α=.01, an a priori power analysis conducted using G*Power 3.1 [36] indicated a total sample size of 115 was required to achieve a power of 80%. A greater number of participants (ie, n=166) were recruited to account for potential exclusions. Participants who did not complete both components of the study (n=9) were excluded from the analysis; this left 79 participants who had undertaken the independent task prior to visiting the laboratory and 78 participants who completed the independent task following their laboratory visit. Further, participants were excluded if they did not meet SST inclusion criteria when undertaking the task in the laboratory (n=14) or independently (n=24; Figure 3).

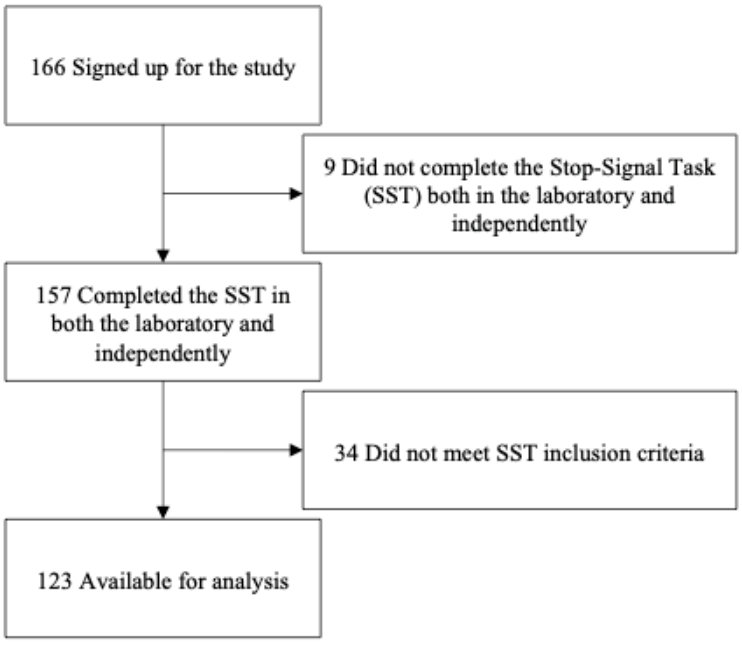
Regarding the SST, given possible variations in timing related to the operating system and browser being used (particularly during the independent component), the program was designed

XSL•FO
RenderX

to capture timing information from the internal timing device, or real-time clock (RTC), of each computer. RTCs are known to be highly accurate [37]. Meta-SSD thus refers to RTC-derived SSD, as opposed to programmed SSD; average SSRT was calculated as mean meta-SSD subtracted from the mean go RT (also timed via the RTC) [21]. There were very strong correlations between laboratory-based meta-SDD (mean 250.23, SD 72.64, range 132.33-484.25 milliseconds) and programmed SSD (mean 249.32, SD 72.65, range 131.25-483.33 milliseconds) (r>0.99, P<.001) and between independent meta-SSD (mean 223.11, SD 62.86, range 94.45-438.52 milliseconds) and programmed SSD (mean 219.06, SD 62.82, range 92.71-437.50 milliseconds) (r>0.99, P<.001).

The Kolmogorov-Smirnov test suggested that both independent (P=.20) and laboratory-based (P=.20) SSRTs were normally distributed. Independent t tests were conducted to determine if there were any differences between participants who did or did not meet the SST inclusion criteria. Dependent samples t tests were used to consider differences between laboratory-based and independent SST variables. Where multiple t tests were employed, a critical P value of .005 was adopted to control for multiple comparisons. Effect sizes were computed for t tests using Cohen d and were interpreted in accordance with Cohen guidelines: small effect=0.20, medium effect=0.50, and large effect=0.80 [38]. Bayesian paired samples t testing was additionally conducted to determine the probability of the alternative hypothesis [39]. We adopted the default priors as set by JASP for the Bayesian analyses. In JASP, the prior distribution is defined by a Cauchy distribution centered on zero with a width or scale of 0.707 for t tests. Results are presented in terms of Bayes factor BF10, which represents the probability of the observed data given the alternative hypothesis [40]. Bayes factors greater than 1 provide evidence for the alternative hypothesis: values of 1-3 imply anecdotal evidence, values of 3-10 imply moderate evidence, values of 10-30 imply strong evidence, and values of >30 imply very strong evidence [41]. Bayes factors less than 1 provide evidence for the null hypothesis: values of 0.33-1 imply anecdotal evidence, values of 0.10-0.33 imply moderate evidence, values of 0.03-0.10 imply strong evidence, and values of <0.03 imply very strong evidence [41]. Data files are available on the Open Science Framework [42].

Figure 3. Study Participation Flowchart.



Ethics Approval

This study was performed in line with the principles of the Declaration of Helsinki. The University of Melbourne Human Ethics Committee approved this study (1954220). All participants provided informed consent. In doing so, they acknowledged reading a plain-language statement that explains that aggregated group level data from this study may be published or presented at conferences.

Results

After exclusions, data from 123 participants (mean age 19.73, SD 1.97, range 18-36 years; 111/123, 90% female) were available for analysis. The Fisher exact test revealed that neither males nor females were excluded to a significantly greater extent in either the laboratory-based (P=.66) or independent (P=.48) conditions. There were no significant differences between participant samples that did or did not meet the inclusion criteria for the laboratory-based SST with regard to age (t_{155}=–0.45; P=.65; r=0.04), years of education (t_{155}=–0.35; P=.73; r=0.03), AUDIT score (t_{155}=0.31; P=.76; r=0.03), or ASSIST score (t_{155}=0.10; P=.92; r=0.01). There were also no significant differences between participant samples that did or did not meet the inclusion criteria for the independent SST on age (t_{155}=0.73; P=.47; r=0.06), years of education (t_{155}=–1.20; P=.23; r=0.10), AUDIT score (t_{155}=0.47; P=.64; r=0.04), or ASSIST score ($t_{24.20}$=–1.36; P=.19; r=0.27). There was no significant correlation between the SSRT—whether derived from the

XSL•FO
RenderX

laboratory-based or independent SST—and age, years of education, AUDIT score, or ASSIST score (Table 1).

Average time between SST testing was 3.72 (SD 2.86) days. Table 2 shows the mean (SD) values of laboratory-based and independent SST variables, differences between variables, and the results of Bayesian analyses. Overall, there was a significant positive correlation between laboratory-based and independent SSRT (r=0.48; P<.001; 95% CI 0.33-0.61). In the group that completed the independent task prior to visiting the laboratory (n=63), there was a slightly stronger association between independent (mean 267.45, SD 35.84 milliseconds) and laboratory-based (mean 239.54, SD 32.37 milliseconds) SSRT (r=0.53; P<.001; 95% CI 0.39-0.65). There was no significant difference between independent (mean 97.54, SD 1.93) and laboratory-based (mean 97.95, SD 2.01) go accuracy (t_{62}=–1.57; P=.12) in this group. The relationship between laboratory-based and independent SSRT was marginally less strong in the group (n=60) that completed the independent task (mean 257.21, SD 36.59 milliseconds) after visiting the laboratory (mean 241.56, SD 39.03 milliseconds; r=0.45; P<.001; 95% CI 0.30-0.58). There was a significant difference between independent (mean 96.68, SD 2.62) and laboratory-based (mean 97.96, SD 1.80) go accuracy (t_{59}=–4.23; P<.001) in this group.

Table 1. Demographic statistics and Pearson correlation coefficients with laboratory-based and independent stop-signal reaction times (n=123).

Characteristics	Value, mean (SD)	Correlation with SSRT[a]			
		Laboratory-based stop-signal task, *r*	*P* value	Independent stop-signal task, *r*	*P* value
Age (years)	19.73 (1.97)	0.09	.34	0.07	.45
Education (years)	13.50 (1.07)	0.01	.94	0.06	.54
Alcohol Use Disorders Identification Test score	4.60 (4.50)	0.15	.10	0.17	.07
Alcohol, Smoking and Substance Involvement Screen Test score	4.07 (8.68)	0.17	.05	0.16	.08

[a]SSRT: stop-signal reaction time (Go reaction time – meta–stop signal delay).

Table 2. Laboratory-based and independent Stop-Signal Task (SST) variables plus differences between variables.

Variables	SST, mean (SD)		*t* test (*df*)	*P* value	95% CI	Cohen *d*	Bayes factor
	Laboratory-based	Independent					
Go accuracy	97.96 (1.90)	97.12 (2.33)	4.12 (122)	<.001	0.43 to 1.24	0.37	242.35[a]
Go reaction time (milliseconds)	490.75 (62.71)	485.56 (60.88)	1.00 (122)	.32	–5.13 to 15.52	0.09	0.16
Go omissions	0.61 (0.99)	0.67 (1.35)	–0.44 (120)	.66	–0.29 to 0.18	0.04	0.11
Go errors	1.43 (1.59)	2.22 (1.97)	–4.24 (120)	<.001	–1.13 to –0.41	0.39	362.27[a]
Go errors reaction time (milliseconds)	431.43 (87.46)	413.44 (82.88)	2.48 (69)	.02	6.64 to 61.05	0.30	2.26[a]
Intraindividual SD	83.03 (26.09)	86.20 (22.60)	–1.27 (122)	.21	–8.10 to 1.76	0.12	0.22
Stop accuracy	50.10 (2.32)	49.61 (2.12)	2.23 (122)	.03	0.05 to 0.93	0.20	1.08
Failed (correct key) stop reaction time (milliseconds)	452.04 (53.27)	446.04 (51.81)	1.26 (120)	.21	–3.22 to 14.56	0.12	0.22
Failed (incorrect key) stop reaction time (milliseconds)	379.58 (58.23)	339.29 (80.89)	2.01 (24)	.06	–1.24 to 99.94	0.40	1.18
Meta stop-signal delay (as timed by each participant's computer; milliseconds)	250.23 (72.64)	223.11 (62.86)	4.42 (122)	<.001	14.98 to 39.26	0.40	718.50[a]
Stop-signal reaction time[b] (milliseconds)	240.53 (35.64)	262.45 (36.43)	–6.61 (122)	<.001	–28.48 to –15.36	0.60	9050000[a]

[a]Evidence for the alternative hypothesis.

[b]Stop-signal reaction time = Go reaction time – meta–stop-signal delay.

XSL•FO
RenderX

Discussion

Principal Findings

In this study, we sought to ascertain whether performance on a web-based version of the SST differed as a function of the testing environment. Using a counterbalanced design, we investigated if there were differences between variables derived from the task when it was undertaken independently—that is, outside the laboratory, on any computer, and in the absence of researchers—versus when it was performed under laboratory conditions. In keeping with our hypothesis, we found that there was a positive correlation between independent and laboratory-based SSRT. Indeed, this relationship was stronger when the independent SST was completed prior to the laboratory-based measure. Correlations were largely consistent with SSRT test-retest reliabilities reported in other (laboratory-based) studies involving healthy participants (r=0.43-0.65) [15,43]. As expected, the independent SST yielded significantly lower go accuracy, increased go errors, and longer SSRTs. Regardless of condition, however, there was no difference in go RT, go omissions, stop accuracy, or intraindividual variability. Bayesian analyses provided very strong evidence in support of the alternative hypothesis in the case of go accuracy, go errors, and SSRT; there was moderate evidence in support of the null hypothesis in the case of go omissions and intraindividual variability. Data could be consistent with either the alternative or null hypothesis in the case of stop accuracy.

Results were largely consistent with an emerging body of evidence examining how the testing environment impacts performance on web-based cognitive tasks. In a study that compared the independent versus fully supervised performance of older adults on web-based tasks assessing attention, memory, and elements of executive function, correlations ranged r=0.42-0.64 [32]. In younger participants, correlations of r=0.40-0.73 have been reported between test results obtained in person and digitally in assessments of recognition, memory, planning, and attention [31]. As with this study, researchers have noted that these correlations accord with test-retest values reported in the psychometric literature.

Our findings lend some support to the efficacy of employing an independent web-based SST to assess response inhibition in the healthy population. Traditionally, SST data are collected in the laboratory using the same stand-alone computer across all participants so as to reduce variability related to setting and computer hardware or software [23]. This means, however, that participants must visit a research laboratory to take part in studies using this task, and, often, they must be assessed individually. As a consequence, sample size and diversity may be limited. This, in turn, impacts statistical power and the generalizability of the findings. Moreover, these issues may be amplified given the advent of the COVID-19 pandemic and associated government-mandated restrictions. A web-based version of the SST, which ensures that the task is accessible to virtually any person at any location, may minimize these limitations.

Web-based survey–based psychological assessment has been recognized as a cost-effective, efficient, and psychometrically sound means of recruiting large, diverse samples [26,28,44-48]. Web-based versions of cognitive tasks may similarly allow for greater participation in cognitive psychological research. For instance, they will likely enable a greater number of persons located in rural and remote communities to participate in cognitive studies. In terms of SUD-focused research, this may be especially useful in a nation such as Australia, where people living in rural and remote communities consume alcohol at harmful levels or use illicit drugs to a greater extent than those living in urban locations [49]. In fact, the web-based SST was recently used in a study focusing on at-risk drinking and vulnerability for transition to dependence [50]. Researchers secured a large sample (N=814) that was representative of the wider Australian population in terms of country of birth and first language; importantly, more than 10% of the sample heralded from rural or remote regions [50]. Web-based cognitive tasks might additionally facilitate easier access to other hard-to-reach samples—such as older individuals, persons living with mobility issues, culturally diverse groups, or those in treatment [46,47,51]—that tend to be underrepresented in psychological research [52-56].

As web-based studies have been found to foster an increased sense of anonymity and confidentiality among participants, potentially decreasing social response bias and increasing the accuracy of data [47], participants subject to discrimination or stigma might be more willing to take part in cognitive research when protocols are entirely web-based. Where sensitive information pertaining to drug and alcohol use is collected, this is likely to be particularly useful [57]. Finally, web-based cognitive tasks would make collaboration between researchers located in different geographical regions more streamlined, providing participants in any country with ready access to the same protocols.

Limitations and Further Research

While our results are promising, the independent condition was characterized by a greater number of exclusions (24/157, 15%) than the laboratory-based condition (14/157, 9%). This may have been due to timing and accuracy offsets related to participant concentration or hardware or software variability in uncontrolled testing environments. Nonetheless, exclusion rates were consistent with those cited in other (laboratory-based) SST studies (4%-17%) [17,43,58]. Interestingly, the association between independent and laboratory-based SSRT was stronger when participants completed the task on their own devices prior to visiting the laboratory. This accords with findings in other similar studies examining comparability between independent and laboratory-based cognitive tests [30]. It may be that when undertaking the SST for the first time, participants find laboratory-based testing relatively more stressful—owing to increased researcher supervision, for instance—than when completing the task independently. This may induce a greater degree of task fatigue such that performance is attenuated when it is undertaken the second time. This accords with the significant difference between laboratory-based and independent go accuracy in the group that completed the task in the

XSL•FO
RenderX

laboratory prior to undertaking it independently. It would be interesting to examine this proposition further in future.

Regardless, participants exhibited reduced inhibitory control when undertaking the independent SST. This may be owing to the uncontrolled nature of the testing environment in this condition. Response inhibition performance is diminished when attention is compromised, such as when fatigued or under high working memory load [59,60]. To minimize the impact of potential environmental distractors during the independent SST, instructions to participants should include explicit directives to undertake the task at quiet locations. Underperformance owing to nonserious testing attitudes might be an issue [61]. Although this is likely to have been randomized across the whole sample, it would be interesting to determine whether more (or less) impulsive individuals are more susceptible to this phenomenon in future studies. Variability might also emerge owing to the use of different computers and browsers, as well as internet speed [61]. While participants used a surprisingly constrained range of operating systems and browsers when undertaking the task independently (73% used Mac operating systems while Google Chrome or Safari was the browser of choice for 94%), these differed somewhat from those used in the laboratory (Windows, 100%; Google Chrome, 100%). Nonetheless, it is important to remember that behavioral measures of impulsivity are designed to capture transient fluctuations in impulsivity, and variations in performance are expected arise in response to various stimuli and environmental conditions [1,62,63]. As such, SSRT data might be expected to vary between sessions.

Several other limitations should be noted. As the study was advertised via a research participation webpage hosted by the Melbourne School of Psychological Sciences, most participants were undergraduate psychology students. The majority were thus female, and age was positively skewed. Although we can make preliminary assertions regarding the validity of the web-based SST, further research is required to determine whether it satisfactorily reveals changes in response inhibition across the lifespan or if it detects response inhibition deficits in clinical settings. Future studies could consider examining how individuals diagnosed with ADHD or SUDs, for instance, perform on the task as a function of environment or as compared to more traditional forms of the SST. Promisingly, however, the SSRT values reported in this study were consistent with those reported in both a recent meta-analysis [64] and in a psychometric study involving only healthy individuals [65].

Conclusions

While further testing is required to determine the association between independent and laboratory-based SST variables among individuals diagnosed with clinical conditions, our findings nevertheless suggest that response inhibition can be measured by a web-based SST undertaken outside the laboratory, on any computer, and in the absence of any researcher. The task could, in future, be used as part of a wider battery of assessments conducted entirely digitally and might thus assist in contending with methodological limitations pertaining to sample size and diversity.

Acknowledgments

This research was supported by an Australian National Health and Medical Research Council grant (1050766) and an Australian Research Council fellowship (FT110100088).

Data and Materials Availability

The data set analyzed in this study is available in the Open Science Framework repository [42]. The SST can be accessed on the Hester Lab website [35].

Authors' Contributions

All authors made substantial contribution to the conception, design, gathering, analysis and interpretation of data, and they have all contributed to the writing and intellectual content of the manuscript.

Conflicts of Interest

None declared.

References

1. Cyders MA, Coskunpinar A. Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? Clin Psychol Rev 2011 Aug;31(6):965-982. [doi: 10.1016/j.cpr.2011.06.001] [Medline: 21733491]
2. Friedman NP, Miyake A. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. Cortex 2017 Jan;86:186-204 [FREE Full text] [doi: 10.1016/j.cortex.2016.04.023] [Medline: 27251123]
3. Herman AM, Duka T. Facets of impulsivity and alcohol use: What role do emotions play? Neurosci Biobehav Rev 2019 Nov;106:202-216. [doi: 10.1016/j.neubiorev.2018.08.011] [Medline: 30343823]
4. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: A latent variable analysis. Cogn Psychol 2000 Aug;41(1):49-100. [doi: 10.1006/cogp.1999.0734] [Medline: 10945922]

XSL•FO
RenderX

5. Diamond A. Executive functions. Annu Rev Psychol 2013;64:135-168 [FREE Full text] [doi: 10.1146/annurev-psych-113011-143750] [Medline: 23020641]
6. Criaud M, Boulinguez P. Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. Neurosci Biobehav Rev 2013 Jan;37(1):11-23. [doi: 10.1016/j.neubiorev.2012.11.003] [Medline: 23164813]
7. Colzato L, Hertsig G, van den Wildenberg W, Hommel B. Estrogen modulates inhibitory control in healthy human females: Evidence from the stop-signal paradigm. Neuroscience 2010 May 19;167(3):709-715. [doi: 10.1016/j.neuroscience.2010.02.029] [Medline: 20219635]
8. Howard SJ, Johnson J, Pascual-Leone J. Clarifying inhibitory control: Diversity and development of attentional inhibition. Cognitive Development 2014 Jul;31:1-21. [doi: 10.1016/j.cogdev.2014.03.001]
9. Reynolds B, Ortengren A, Richards JB, de Wit H. Dimensions of impulsive behavior: Personality and behavioral measures. Personality and Individual Differences 2006 Jan;40(2):305-315. [doi: 10.1016/j.paid.2005.03.024]
10. Albinet CT, Boucard G, Bouquet CA, Audiffren M. Processing speed and executive functions in cognitive aging: How to disentangle their mutual relationship? Brain Cogn 2012 Jun;79(1):1-11. [doi: 10.1016/j.bandc.2012.02.001] [Medline: 22387275]
11. Lipszyc J, Schachar R. Inhibitory control and psychopathology: A meta-analysis of studies using the stop signal task. J Int Neuropsychol Soc 2010 Nov;16(6):1064-1076. [doi: 10.1017/S1355617710000895] [Medline: 20719043]
12. Lijffijt M, Kenemans JL, Verbaten MN, van Engeland H. A meta-analytic review of stopping performance in attention-deficit/hyperactivity disorder: Deficient inhibitory motor control? J Abnorm Psychol 2005 May;114(2):216-222. [doi: 10.1037/0021-843X.114.2.216] [Medline: 15869352]
13. Stevens L, Verdejo-García A, Goudriaan AE, Roeyers H, Dom G, Vanderplasschen W. Impulsivity as a vulnerability factor for poor addiction treatment outcomes: A review of neurocognitive findings among individuals with substance use disorders. J Subst Abuse Treat 2014 Jul;47(1):58-72. [doi: 10.1016/j.jsat.2014.01.008] [Medline: 24629886]
14. Verdejo-García A, Lawrence AJ, Clark L. Impulsivity as a vulnerability marker for substance-use disorders: Review of findings from high-risk research, problem gamblers and genetic association studies. Neurosci Biobehav Rev 2008;32(4):777-810. [doi: 10.1016/j.neubiorev.2007.11.003] [Medline: 18295884]
15. Weafer J, Baggott MJ, de Wit H. Test-retest reliability of behavioral measures of impulsive choice, impulsive action, and inattention. Exp Clin Psychopharmacol 2013 Dec;21(6):475-481 [FREE Full text] [doi: 10.1037/a0033659] [Medline: 24099351]
16. Wöstmann NM, Aichert DS, Costa A, Rubia K, Möller HJ, Ettinger U. Reliability and plasticity of response inhibition and interference control. Brain Cogn 2013 Feb;81(1):82-94. [doi: 10.1016/j.bandc.2012.09.010] [Medline: 23174432]
17. Congdon E, Mumford JA, Cohen JR, Galvan A, Canli T, Poldrack RA. Measurement and reliability of response inhibition. Front Psychol 2012;3:37 [FREE Full text] [doi: 10.3389/fpsyg.2012.00037] [Medline: 22363308]
18. Dalley JW, Robbins TW. Fractionating impulsivity: Neuropsychiatric implications. Nat Rev Neurosci 2017 Feb 17;18(3):158-171. [doi: 10.1038/nrn.2017.8] [Medline: 28209979]
19. Herman AM, Critchley HD, Duka T. The role of emotions and physiological arousal in modulating impulsive behaviour. Biol Psychol 2018 Mar;133:30-43. [doi: 10.1016/j.biopsycho.2018.01.014] [Medline: 29391179]
20. Logan G, Cowan WB. On the ability to inhibit thought and action: A theory of an act of control. Psychol Rev 1984 Jan;91(3):295-327. [doi: 10.1037/0033-295X.91.3.295]
21. Verbruggen F, Aron AR, Band GP, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife 2019 Apr 29;8 [FREE Full text] [doi: 10.7554/eLife.46323] [Medline: 31033438]
22. Logan GD, Schachar RJ, Tannock R. Impulsivity and inhibitory control. Psychol Sci 1997 Jan;8(1):60-64. [doi: 10.1111/j.1467-9280.1997.tb00545.x]
23. Verbruggen F, Logan GD, Stevens MA. STOP-IT: Windows executable software for the stop-signal paradigm. Behav Res Methods 2008 May;40(2):479-483. [doi: 10.3758/brm.40.2.479] [Medline: 18522058]
24. Logan GD, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: A model and a method. J Exp Psychol Hum Percept Perform 1984 Apr;10(2):276-291. [doi: 10.1037//0096-1523.10.2.276] [Medline: 6232345]
25. Gentili C, Cristea IA. Challenges and opportunities for human behavior research in the coronavirus disease (COVID-19) pandemic. Front Psychol 2020;11:1786 [FREE Full text] [doi: 10.3389/fpsyg.2020.01786] [Medline: 32754106]
26. Sassenberg K, Ditrich L. Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. Adv Meth Pract Psychol Sci 2019 Apr 12;2(2):107-114. [doi: 10.1177/2515245919838781]
27. Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012 Oct;19(5):847-857. [doi: 10.3758/s13423-012-0296-9] [Medline: 22829343]

XSL•FO
RenderX

28. Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. Am Psychol 2004;59(2):93-104. [doi: 10.1037/0003-066X.59.2.93] [Medline: 14992636]
29. Semmelmann K, Weigelt S. Online psychophysics: Reaction time effects in cognitive experiments. Behav Res Methods 2017 Aug;49(4):1241-1260. [doi: 10.3758/s13428-016-0783-4] [Medline: 27496171]
30. Cyr A, Romero K, Galin-Corini L. Web-based cognitive testing of older adults in person versus at home: within-subjects comparison study. JMIR Aging 2021 Feb 01;4(1):e23384 [FREE Full text] [doi: 10.2196/23384] [Medline: 33522972]
31. Backx R, Skirrow C, Dente P, Barnett JH, Cormack FK. Comparing web-based and lab-based cognitive assessment using the Cambridge Neuropsychological Test Automated Battery: A within-subjects counterbalanced study. J Med Internet Res 2020 Aug 04;22(8):e16792 [FREE Full text] [doi: 10.2196/16792] [Medline: 32749999]
32. Assmann KE, Bailet M, Lecoffre AC, Galan P, Hercberg S, Amieva H, et al. Comparison between a self-administered and supervised version of a web-based cognitive test battery: Results from the NutriNet-Santé cohort study. J Med Internet Res 2016 Apr 05;18(4):e68 [FREE Full text] [doi: 10.2196/jmir.4862] [Medline: 27049114]
33. 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 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
34. 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] [Medline: 12199834]
35. Stop-Signal Task Internet. Hesterlab. 2015. URL: http://hesterlab.org/tasks/t1/ssd_task.html [accessed 2022-05-10]
36. Faul F, Erdfelder E, Buchner A, Lang A. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav Res Methods 2009 Nov;41(4):1149-1160. [doi: 10.3758/BRM.41.4.1149] [Medline: 19897823]
37. Marouani H, Dagenais MR. Internal clock drift estimation in computer clusters. J Comput Sys Netw Commun 2008;2008:1-7. [doi: 10.1155/2008/583162]
38. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd edition). Hillsdale, NY: Lawrence Erlbaum Associates; 1988.
39. Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 2009 Apr;16(2):225-237. [doi: 10.3758/PBR.16.2.225] [Medline: 19293088]
40. Wagenmakers E, Love J, Marsman M, Jamil T, Ly A, Verhagen J, et al. Bayesian inference for psychology. Part II: Example applications with JASP. Psychon Bull Rev 2018 Feb;25(1):58-76 [FREE Full text] [doi: 10.3758/s13423-017-1323-7] [Medline: 28685272]
41. Lee M, Wagenmakers EJ. Bayesian Cognitive Modeling: A Practical Course. Cambridge: Cambridge University Press; 2013.
42. Poulton A. Measuring response inhibition online: Comparing independent and laboratory-based Stop-Signal Task performance. 2021. URL: https://osf.io/754m2/ [accessed 2022-05-10]
43. Hedge C, Powell G, Sumner P. The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behav Res Methods 2018 Jun;50(3):1166-1186 [FREE Full text] [doi: 10.3758/s13428-017-0935-1] [Medline: 28726177]
44. Gosling SD, Mason W. Internet research in psychology. Annu Rev Psychol 2015 Jan 03;66:877-902. [doi: 10.1146/annurev-psych-010814-015321] [Medline: 25251483]
45. Epstein J, Klinkenberg W, Wiley D, McKinley L. Insuring sample equivalence across internet and paper-and-pencil assessments. Comput Hum Behav 2001 May;17(3):339-346. [doi: 10.1016/s0747-5632(01)00002-4]
46. Weigold A, Weigold IK, Russell EJ. Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. Psychol Methods 2013 Mar;18(1):53-70. [doi: 10.1037/a0031607] [Medline: 23477606]
47. Cantrell MA, Lupinacci P. Methodological issues in online data collection. J Adv Nurs 2007 Dec;60(5):544-549. [doi: 10.1111/j.1365-2648.2007.04448.x] [Medline: 17973718]
48. Reimers S. The BBC internet study: General methodology. Arch Sex Behav 2007 Apr;36(2):147-161. [doi: 10.1007/s10508-006-9143-2] [Medline: 17380377]
49. Alcohol and other drug use in regional and remote Australia: Consumption, harms and access to treatment 2016–17. Australian Institute of Health and Welfare. Australian Government. 2019. URL: https://www.aihw.gov.au/reports/alcohol-other-drug-treatment-services/alcohol-other-drug-use-regional-remote-2016-17/data [accessed 2022-05-10]
50. Poulton A, Eastwood O, Bruns LR, Sinnott RO, Hester R. Addressing methodological issues in a study of impulsivity and vulnerability for transition to alcohol use disorder. Alcohol Clin Exp Res 2022 Feb;46(2):262-276. [doi: 10.1111/acer.14755] [Medline: 34859438]
51. Naglieri JA, Drasgow F, Schmit M, Handler L, Prifitera A, Margolis A, et al. Psychological testing on the Internet: new problems, old issues. Am Psychol 2004 Apr;59(3):150-162. [doi: 10.1037/0003-066X.59.3.150] [Medline: 15222858]
52. Cundiff JL. Is mainstream psychological research “womanless” and “raceless”? An updated analysis. Sex Roles 2012 Mar 20;67(3-4):158-173. [doi: 10.1007/s11199-012-0141-7]

XSL•FO
RenderX

53. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant: Solution for recruiting hidden, hard-to-reach, and vulnerable populations. Int J Qual Methods 2015 Dec 17;14(5):160940691562142. [doi: 10.1177/1609406915621420]
54. Thalmayer AG, Toscanelli C, Arnett JJ. The neglected 95% revisited: Is American psychology becoming less American? Am Psychol 2021 Jan;76(1):116-129. [doi: 10.1037/amp0000622] [Medline: 32271027]
55. Richmond AS, Broussard KA, Sterns JL, Sanders KK, Shardy JC. Who are we studying? Sample diversity in teaching of psychology research. Teach Psychol 2015 May 22;42(3):218-226. [doi: 10.1177/0098628315587619]
56. Thurman WA, Harrison TC. Reaching the "hard-to-reach": Recruitment of rural-dwelling adults with disabilities. J Transcult Nurs 2020 Mar;31(2):171-177. [doi: 10.1177/1043659619856667] [Medline: 31204586]
57. Bates SC, Cox JM. The impact of computer versus paper–pencil survey, and individual versus group administration, on self-reports of sensitive behaviors. Comput Hum Behav 2008 May;24(3):903-916. [doi: 10.1016/j.chb.2007.02.021]
58. Kräplin A, Scherbaum S, Bühringer G, Goschke T. Retest reliabilities of decision-making and cognitive control measures in addictive disorders. SUCHT 2016 Aug;62(4):191-202. [doi: 10.1024/0939-5911/a000430]
59. Nyberg L, Brocki K, Tillman C, Bohlin G. The proposed interaction between working memory and inhibition. Eur J Cogn Psychol 2009 Feb;21(1):84-111. [doi: 10.1080/09541440701862133]
60. Demos K, Hart C, Sweet L, Mailloux K, Trautvetter J, Williams S, et al. Partial sleep deprivation impacts impulsive action but not impulsive decision-making. Physiol Behav 2016 Oct 01;164(Pt A):214-219 [FREE Full text] [doi: 10.1016/j.physbeh.2016.06.003] [Medline: 27267950]
61. Feenstra H, Vermeulen I, Murre J, Schagen S. Online cognition: Factors facilitating reliable online neuropsychological test results. Clin Neuropsychol 2017 Jan;31(1):59-84. [doi: 10.1080/13854046.2016.1190405] [Medline: 27266677]
62. Gullo MJ, Potenza MN. Impulsivity: Mechanisms, moderators and implications for addictive behaviors. Addict Behav 2014 Nov;39(11):1543-1546. [doi: 10.1016/j.addbeh.2014.06.005] [Medline: 25042111]
63. Dougherty DM, Mathias CW, Marsh DM, Jagar AA. Laboratory behavioral measures of impulsivity. Behav Res Methods 2005 Feb;37(1):82-90. [doi: 10.3758/bf03206401] [Medline: 16097347]
64. Smith JL, Mattick RP, Jamadar SD, Iredale JM. Deficits in behavioural inhibition in substance abuse and addiction: A meta-analysis. Drug Alcohol Depend 2014 Dec 01;145:1-33. [doi: 10.1016/j.drugalcdep.2014.08.009] [Medline: 25195081]
65. Caswell A, Bond R, Duka T, Morgan M. Further evidence of the heterogeneous nature of impulsivity. Pers Individ Dif 2015 Apr;76:68-74 [FREE Full text] [doi: 10.1016/j.paid.2014.11.059] [Medline: 25844002]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ASSIST: Alcohol, Smoking and Substance Involvement Screen Test
AUDIT: Alcohol Use Disorders Identification Test
ITI: intertrial interval
RT: reaction time
RTC: real-time clock
SSD: stop-signal delay
SSRT: stop-signal reaction time
SST: Stop-Signal Task
SUD: substance use disorder

Edited by R Kukafka; submitted 15.08.21; peer-reviewed by A Staudt, H Li; comments to author 13.12.21; revised version received 03.02.22; accepted 23.02.22; published 30.05.22.

Please cite as:
Poulton A, Chen LPE, Dali G, Fox M, Hester R
Web-Based Independent Versus Laboratory-Based Stop-Signal Task Performance: Within-Subjects Counterbalanced Comparison Study
J Med Internet Res 2022;24(5):e32922
URL: https://www.jmir.org/2022/5/e32922
doi:10.2196/32922
PMID:35635745

©Antoinette Poulton, Li Peng Evelyn Chen, Gezelle Dali, Michael Fox, Robert Hester. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.05.2022. 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,

XSL•FO
RenderX

and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

The Development and Evaluation of an e-Learning Course That Promotes Digital Health Literacy in School-age Children: Pre-Post Measurement Study

Lars König[1], Dr rer nat; Eugenia Marbach-Breitrück[1], Dr rer nat; Anne Engler[1], MSc; Ralf Suhr[1], Dr med

Stiftung Gesundheitswissen, Berlin, Germany

Corresponding Author:
Lars König, Dr rer nat
Stiftung Gesundheitswissen
Friedrichstr 134
Berlin, 10117
Germany
Phone: 49 30419549262
Email: lars.koenig@stiftung-gesundheitswissen.de

Abstract

Background: Experts agree that the promotion of (digital) health literacy should be an integral part of the school curriculum. However, promoting (digital) health literacy within the German school system is difficult because (digital) health education is not a mandatory school subject in all the German states. Therefore, experts suggest that (digital) health literacy could be addressed as part of the mandatory framework for digital education and digital literacy in schools developed by the German Conference on Education Ministries and Cultural Affairs (Kultusministerkonferenz).

Objective: The goal of this study was to evaluate a newly developed e-learning course that was designed to improve (digital) health literacy in school-age children and concurrently to teach skills specified in the mandatory framework for digital education and digital literacy in schools. It was hypothesized that participants' health literacy and digital health literacy levels would be higher after completing the e-learning course than they were before doing the course. Furthermore, it was hypothesized that after completing the e-learning course, participants' subjective and objective knowledge in the domain of (digital) health literacy would be higher than it was before doing the course.

Methods: The pre-post measurement study was conducted online. After participants (N=323) gave their informed consent to participate in the study, they provided demographic information and answered all measures (premeasurement). Following this, participants had 7 days to complete the e-learning course. After finishing the e-learning course, participants answered all the measures again (postmeasurement).

Results: To test the hypotheses, Bayesian paired samples *t* tests (1-sided) were conducted. After completing the e-learning course, participants showed higher health literacy levels. Specifically, they showed higher competency levels in the domains of theoretical knowledge (Bayes factor $[BF]_{-0}$=676,000; δ=–0.316), practical knowledge (BF_{-0}=92,300; δ=–0.294), critical thinking (BF_{-0}=7.42e+13; δ=–0.482), self-awareness (BF_{-0}=11,500,000; δ=–0.345), and citizenship (BF_{-0}=266,000; δ=–0.306). Furthermore, participants achieved higher digital health literacy levels. Specifically, they achieved higher competency levels in the domains of information searching (BF_{-0}=2.339; δ=–0.135), evaluating reliability (BF_{-0}=2.03e+11; δ=–0.434), and determining relevance (BF_{-0}=316,000; δ=–0.308). Moreover, participants demonstrated higher subjective (BF_{-0}=3.58e+82; δ=–1.515) and objective knowledge (BF_{-0}=3.82e+97; δ=–1.758) in the domain of (digital) health literacy.

Conclusions: The newly designed e-learning course provides an easy way for schools and teachers from all German states to integrate (digital) health literacy education into their school curriculums and lessons. The evaluated course is especially attractive because it was designed to improve (digital) health literacy and at the same time to teach skills specified in the mandatory framework for digital education and digital literacy in schools developed by the German Conference on Education Ministries and Cultural Affairs (Kultusministerkonferenz).

(J Med Internet Res 2022;24(5):e37523) doi:10.2196/37523

XSL•FO
RenderX

KEYWORDS

digital education; digital health literacy; digital literacy; e-learning; health education; health information; health literacy; learning material; school; school-age children

Introduction

Health literacy describes people's ability to evaluate and apply health information in the context of disease prevention and health promotion [1]. Health literacy is one central component of digital health literacy, which refers to people's ability to effectively use health information derived from online and electronic sources [2]. In recent years, health literacy has become a major public concern because studies have shown that it is linked to various health outcomes and behaviors [3-7]. People with low health literacy, for example, show poorer dieting habits, are less physically active, smoke more frequently, have more sick leave days, and rely more heavily on the health care system [7]. Furthermore, digital health literacy has become an increasingly important topic because people regularly turn to the internet when searching for health information [8-12]. In such situations, people are often confronted with inaccurate online health information and therefore need the ability to evaluate the trustworthiness of information sources and the credibility of their information [10,13-18]. Teaching adequate evaluation strategies seems especially important because laypeople often base their trustworthiness judgments on factors like the enthusiasm of an information source and their tone of voice [19].

In Germany, health literacy has decreased since 2014 [5,6]. A representative survey published in 2021 found that 58.8% of the German population had low health literacy levels and 75.8% had low digital health literacy levels [6,7]. When searching the internet for health information, it is especially important to decide whether the information is reliable and whether it is written with commercial interests [20]. However, 82.6% of the German population find it difficult or very difficult to decide whether information is reliable and 82% report it as difficult or very difficult to assess whether information is written with commercial interests [7]. Even university students, who represent a well-educated subgroup of the German population, find it difficult to decide whether online health information is reliable and written with commercial interests [21]. Such results may arise because information evaluation strategies are not well enough embedded within the German school system and even preservice teachers have problems adapting to the digitization of the educational system [22]. Data from the Program for International Student Assessment, for example, show that during their entire school experience, only 54.3% of the students were taught how to decide whether to trust information from the internet. Furthermore, only 48.7% of the students were taught how to detect whether information is subjective or biased, and only 45.2% had the capacity to distinguish facts from opinions [23].

Such results are unfortunate because teaching (digital) health literacy to school-age children has the potential to improve various health outcomes later in life [24,25]. Furthermore, schools seem to be an ideal place to promote (digital) health literacy because they can reach almost all children within a society [26-28]. In line with this argumentation, the World Health Organization argues that health literacy should be an integral part of the school curriculum [29]. Furthermore, a recent concept paper from the World Health Organization Regional Office for Europe specifically stresses the importance of addressing health literacy as well as digital health literacy in schools [30]. For Germany, addressing (digital) health literacy in schools seems especially important because German pupils demonstrate particularly low health literacy levels compared to pupils from other European countries [31]. However, promoting (digital) health literacy in schools is difficult because (digital) health education is not a mandatory school subject in all German states. To address this issue, experts suggest that (digital) health literacy could be addressed as part of the mandatory framework for digital education and digital literacy in schools, which was developed by the German Conference on Education Ministries and Cultural Affairs (Kultusministerkonferenz) [32,33].

In line with these suggestions, the independent, nonprofit foundation Stiftung Gesundheitswissen developed the free e-learning platform Gesundweiser.de to teach (digital) health literacy to school-age children. On the platform, visitors can receive information on the topic of (digital) health literacy. Furthermore, they can take part in a free e-learning course. The course was designed to improve (digital) health literacy and at the same time to teach skills specified in the mandatory framework for digital education and digital literacy in schools [32]. Even though there are various reasons why (digital) health literacy should be taught in schools [34], research has shown that many school-based health interventions end after external funding stops [35]. To facilitate the permanent implementation of an intervention, it is important to design interventions that recognize the specific needs of schools and teachers [36]. Therefore, the e-learning course was designed to be applicable in various types of schools and subjects. Since the course is self-explanatory and no active supervision is required, it is especially suitable as a homework exercise that can be completed within a set amount of time. Within the course, participants learn how they can evaluate health information on the internet. The provided material was created by a multiprofessional team, including health and e-learning specialists, and was derived from professional guidelines (eg, Guideline for the Development of Evidence-based Patient Information) [37]. The aim of this study was to test the following hypotheses:

- Hypothesis 1: Participants' health literacy levels will be higher after completing the e-learning course than they were before completing the e-learning course.
- Hypothesis 2: Participants' digital health literacy levels will be higher after completing the e-learning course than they were before completing the e-learning course.
- Hypothesis 3: Participants' subjective knowledge in the domain of (digital) health literacy will be higher after completing the e-learning course than it was before completing the e-learning course.

XSL•FO
RenderX

- Hypothesis 4: Participants' objective knowledge in the domain of (digital) health literacy will be higher after completing the e-learning course than it was before completing the e-learning course.

Methods

Sample

German pupils between the age of 16 and 20 years were recruited from all German states and different types of schools via an online panel provided by a market research company (SPLENDID RESEARCH GmbH). The German school system is highly complex. Within the 16 German states, various types of schools, subjects, and core curriculums exist. To serve a wide range of pupils, the course was not designed for a specific type of school and core curriculum. Instead, it was designed for pupils between the age of 16 and 20 years. Since 16-year-old pupils typically attend the 10th grade, the course can be used in diverse types of schools with a 10th grade (eg, Hauptschulen, Realschulen, Gesamtschulen, Gymnasien). Furthermore, it can be used in higher grades (eg, Gesamtschulen, Gymnasien) and vocational schools (Berufsschulen) as well. As compensation for participating in the study, participants received a 60€ (exchange rate in January 2022: 1€ ≈ USD 1.1342) online shop voucher. An a priori power analysis using G*Power (University of Düsseldorf) indicated that a total of 272 participants was needed to detect a small-to-medium effect with satisfactory power (specifications: test family = *t* tests; statistical test = means: differences between 2 dependent means [matched pairs]; type of power analysis = a priori: compute required sample size – given α, power, and effect size; tail(s)=1; effects size dz=0.2; α err prob=.05; power [1–β err prob]=0.95) [38]. To compensate for possible participant exclusions and data collection problems, it was decided to oversample slightly.

Ethical Considerations

Before data collection, the study protocol was submitted to the ethics committee of the Berlin Medical Association (Eth-68/21). The ethics committee had no ethical or professional objections to the study protocol.

Procedure

Data collection took place between December 2021 and January 2022. The pre-post measurement study was conducted online using an online platform for data collection provided by a market research company (SPLENDID RESEARCH GmbH). Prior to the study, participants received detailed information about the context, purpose, and procedures of the study. Furthermore, they were informed that they could opt out of the study at any time. After participants gave their informed consent to participate in the study, they provided demographic information and answered all measures (premeasurement). Following this, participants had 7 days to complete the e-learning course. Because the course was not designed for a specific type of school and subject, it can be used in various contexts. Since the course is self-explanatory and no active supervision is required, it is especially suitable as a homework exercise that can be completed within a set amount of time. To simulate such a homework exercise, participants were given 7 days to complete the course. During this 7-day period, participants could use any device to complete the course and they could start and pause the course as often as they liked. Simulating a homework exercise by giving participants the opportunity to complete the course within 7 days has the advantage that it increases ecological validity. However, it also creates methodological disadvantages that will be discussed in the limitations section.

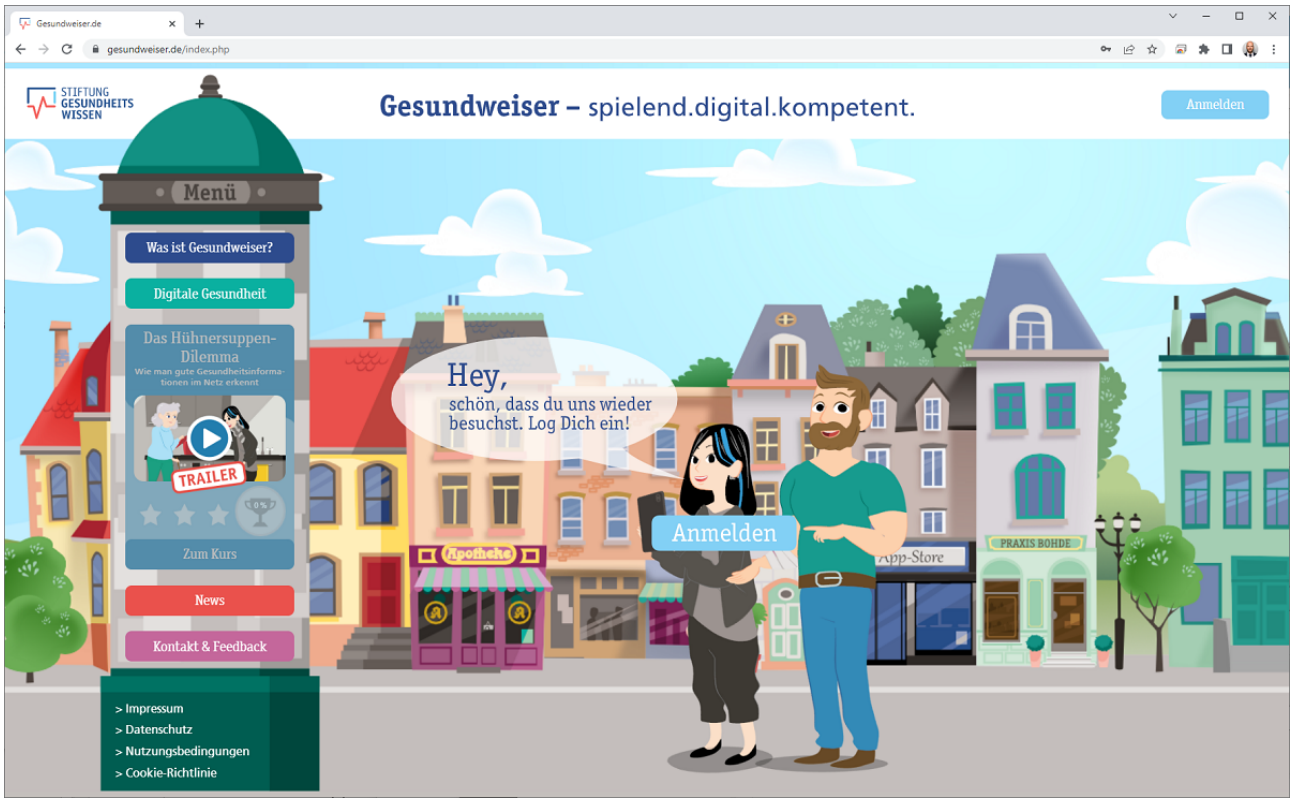
Within the course, participant learned how they could evaluate health information on the internet. The provided material was created by a multiprofessional team, including health and e-learning specialists, and was derived from professional guidelines (eg, Guideline for the Development of Evidence-based Patient Information) [37]. Furthermore, the material was designed to address the competence areas mentioned in the mandatory framework for digital education and digital literacy in schools with a special focus on the competence areas: (1) searching, processing, and storing; (2) problem solving and acting; and (3) analyzing and reflecting [32,33]. The e-learning course consists of 8 mandatory modules, 3 optional modules, and a final test. Internal analyses show that it takes about 4 minutes to complete the shortest module and 18 minutes to complete the longest module. The entire course can be completed in about 2 hours. Table 1 shows the length of the e-learning course modules according to internal analyses. Figure 1 shows a screenshot of the e-learning portal Gesundweiser.de. Figure 2 provides an overview of the e-learning course modules. After finishing the e-learning course, participants answered all the measures again (postmeasurement). At the end of the study, participants were thanked for their participation and received their online shop voucher.

XSL•FO
RenderX

Table 1. Length of the e-learning course modules according to internal analyses.

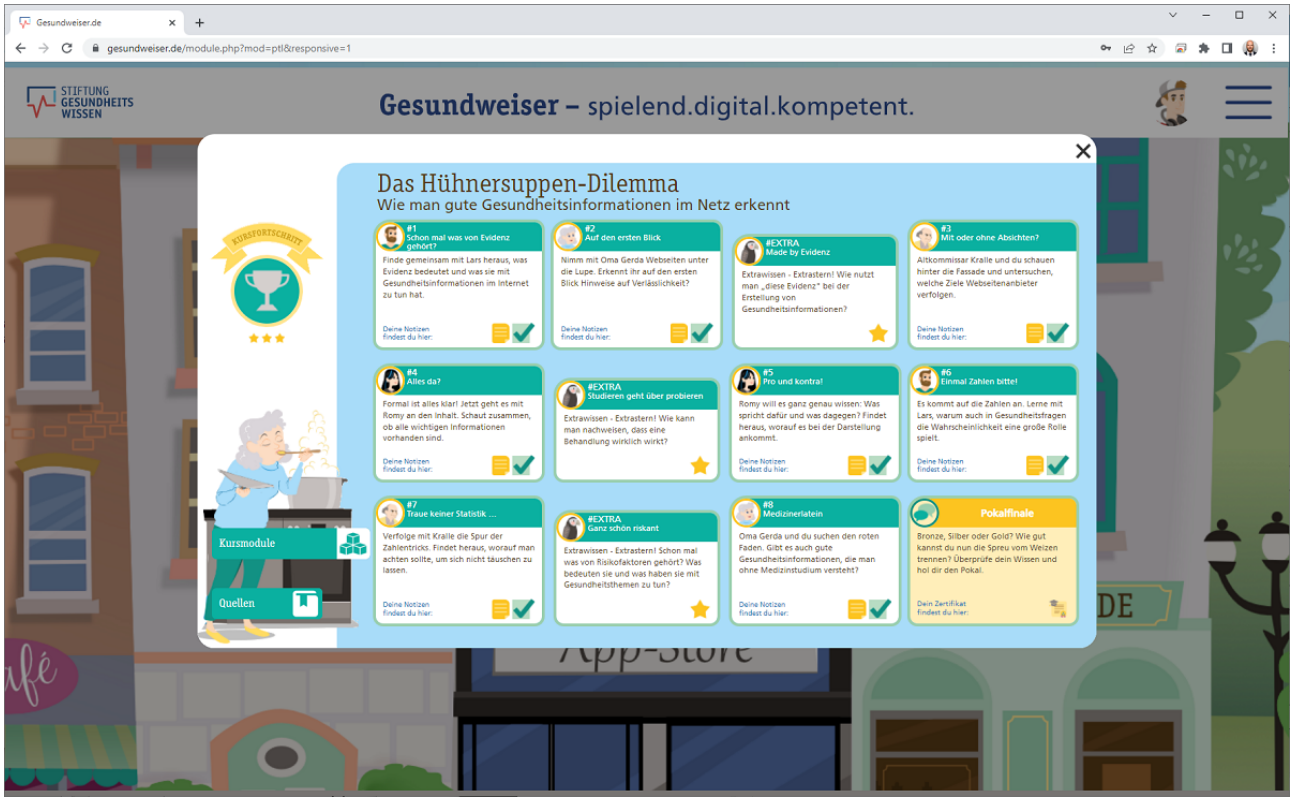
Course module	Length (min)
Mandatory module 1	14
Mandatory module 2	18
Optional module 1	6
Mandatory module 3	6
Mandatory module 4	9
Optional module 2	6
Mandatory module 5	14
Mandatory module 6	12
Mandatory module 7	9
Optional module 3	4
Mandatory module 8	11
Final test	11
Complete course	120

Figure 1. Screenshot of the e-learning platform Gesundweiser.de.



XSL•FO
RenderX

Figure 2. Overview of the e-learning course modules.



Measures

Health Literacy

To assess health literacy, the Health Literacy for School-Age Children Instrument was used [39,40]. This 10-item instrument comprises 5 subscales assessing competencies in the fields of theoretical knowledge (2 items: eg, “I am confident that I have good information about health”), practical knowledge (2 items: eg, “I am confident that when necessary I find health-related information that is easy for me to understand”), critical thinking (2 items: eg, “I am confident that I can usually figure out if some health-related information is right or wrong”), self-awareness (2 items: eg, “I am confident that I can give reasons for choices I make regarding my health”), and citizenship (2 items: eg, “I am confident that I can judge how my own actions affect the surrounding natural environment”). Participants rated all items on scales ranging from 1 (strongly disagree) to 4 (strongly agree). For each subscale, a total score was generated by calculating the mean.

Digital Health Literacy

To assess digital health literacy, 3 of the 7 subscales of the Digital Health Literacy Instrument were used [11,20]. The 3 subscales assessed competencies in the fields of information searching (3 items: eg, “When you search the internet for information on health, how easy or difficult is it for you to make a choice from all the information you find?”), evaluating reliability (3 items: eg, “When you search the internet for information on health, how easy or difficult is it for you to decide whether the information is reliable or not?”), and determining relevance (3 items: eg, “When you search the internet for information on health, how easy or difficult is it for you to decide if the information you found is applicable to you?”). Participants rated all items on scales ranging from 1 (very hard) to 4 (very easy). For each subscale, a total score was generated by calculating the mean.

Subjective (Digital) Health Literacy Knowledge

To assess subjective knowledge in the domain of (digital) health literacy, participants indicated their agreement with 5 items (eg, “I can explain which content-related characteristics are indicative of reliable health information on the internet)”. The items focused on content covered throughout the e-learning course. Participants rated all items on scales ranging from 1 (strongly disagree) to 6 (strongly agree). A total score was generated by calculating the mean.

Objective (Digital) Health Literacy Knowledge

To assess objective knowledge in the domain of (digital) health literacy, participants answered 15 multiple-choice questions, such as “What does the phrase ‘evidence-based’ mean?” with response options (1) reviewed by experts, (2) based on scientific evidence and proof, (3) based on personal views and experiences, and (4) rated as helpful by a certain number of users. The multiple-choice questions focused on content covered throughout the e-learning course. Depending on the multiple-choice question, 1 to 4 of the responses were correct. For each correctly answered multiple-choice question, participants received 1 point. A total score was generated by adding up all points (minimum=0; maximum=15). The original data set contains further variables that have not been described because they exceed the scope of this study.

XSL•FO
RenderX

Statistical Analysis

For all analyses, the statistical software JASP, version 0.16.1 (University of Amsterdam), was used [41]. To test the hypotheses, Bayesian paired-samples *t* tests (1-sided) were conducted with the following specifications: alternative hypothesis (measure 1 < measure 2), Bayes factor (BF_{10}), test (student), missing values (exclude cases per dependent variable), and prior (default Cauchy scale=0.707). These specifications imply that the results will report Bayes factors in favor of the alternative hypotheses (measure 1 < measure 2). Following a commonly used classification scheme, Bayes factors above 1 will be interpreted as anecdotal (1-3), moderate (3-10), strong (10-30), very strong (30-100), or extreme (>100) evidence for the alternative hypothesis compared to the null hypothesis in light of the observed data [42]. Bayes factors below 1 will be interpreted as evidence for the null hypothesis. For all analyses, a Bayes factor robustness check is provided. The robustness check "provides an assessment of the robustness of the Bayes factor under different prior specifications: if the qualitative conclusions do not change across a range of different plausible prior distributions, this indicates that the analysis is relatively robust" [43]. Further information on the interpretation of Bayes factors in medical contexts and nontechnical introductions to Bayesian inference with JASP can be found elsewhere [42-44].

Results

Sample Characteristics

A total of 340 participants completed the study; 17 participants were excluded from data analysis because of data collection problems or because they did not finish the e-learning course. Therefore, the final sample contained 323 (188 females, 132 males, 3 diverse) participants from all German states with an average age of 17.88 (SD 1.22) years. Table 2 shows the sample distribution by state, type of school, and grade.

XSL•FO
RenderX

Table 2. Sample distribution by state, type of school, and grade.

Characteristic	Sample, n
State	
Baden-Württemberg	30
Bayern	41
Berlin	11
Brandenburg	6
Bremen	1
Hamburg	16
Hessen	34
Mecklenburg-Vorpommern	7
Niedersachsen	33
Nordrhein-Westfalen	78
Rheinland-Pfalz	12
Saarland	6
Sachsen	22
Sachsen-Anhalt	7
Schleswig-Holstein	7
Thüringen	12
Type of school	
Hauptschule	1
Realschule	23
Sekundarschule	2
Gesamtschule	37
Gymnasien	134
Berufsschule	78
Berufsfachschule	24
Fachoberschule	15
Other	9
Grade	
8th	0
9th	7
10th	45
11th	57
12th	100
13th	47
Other	67

Findings

Table 3 provides descriptive statistics of the premeasurements and postmeasurements.

XSL•FO
RenderX

Table 3. Descriptive statistics of the premeasurements and postmeasurements[a].

Measure	Mean (SD)	SE	95% Credible interval
Health literacy			
Theoretical knowledge (pre)	2.810 (0.542)	0.030	2.750-2.869
Theoretical knowledge (post)	3.008 (0.510)	0.028	2.952-3.064
Practical knowledge (pre)	3.017 (0.598)	0.033	2.952-3.082
Practical knowledge (post)	3.207 (0.526)	0.029	3.150-3.265
Critical thinking (pre)	2.759 (0.565)	0.031	2.697-2.820
Critical thinking (post)	3.107 (0.587)	0.033	3.043-3.171
Self-awareness (pre)	2.856 (0.583)	0.032	2.792-2.920
Self-awareness (post)	3.088 (0.523)	0.029	3.031-3.146
Citizenship (pre)	2.850 (0.567)	0.032	2.788-2.912
Citizenship (post)	3.053 (0.543)	0.030	2.993-3.112
Digital health literacy			
Information searching (pre)	2.716 (0.607)	0.034	2.650-2.783
Information searching (post)	2.807 (0.566)	0.031	2.745-2.869
Evaluating reliability (pre)	2.522 (0.603)	0.034	2.456-2.588
Evaluating reliability (post)	2.841 (0.605)	0.034	2.775-2.907
Determining relevance (pre)	2.651 (0.583)	0.032	2.587-2.715
Determining relevance (post)	2.853 (0.557)	0.031	2.792-2.914
(Digital) health literacy knowledge			
Subjective (pre)	3.510 (1.068)	0.059	3.393-3.627
Subjective (post)	5.202 (0.708)	0.039	5.124-5.279
Objective (pre)	5.842 (2.870)	0.160	5.528-6.156
Objective (post)	11.022 (2.221)	0.124	10.779-11.265

[a]Health literacy measures ranged from 1 (low score) to 4 (high score); digital health literacy measures ranged from 1 (low score) to 4 (high score); subjective (digital) health literacy knowledge ranged from 1 (low score) to 6 (high score); and objective (digital) health literacy knowledge ranged from 0 (low score) to 15 (high score).

Health Literacy

It was hypothesized that participants' health literacy levels would be higher after completing the e-learning course than they were before completing the e-learning course. The results show extreme evidence for the hypothesis. After completing the e-learning course, participants reported higher competencies in the fields of theoretical knowledge (extreme evidence, BF_{-0}=676,000; δ=–0.316), practical knowledge (extreme evidence, BF_{-0}=92,300; δ=–0.294), critical thinking (extreme evidence, BF_{-0}=7.42e+13; δ=–0.482), self-awareness (extreme evidence, BF_{-0}=11,500,000; δ=–0.345), and citizenship (extreme evidence, BF_{-0}=266,000; δ=–0.306). The corresponding prior and posterior distribution plots, effect sizes, and Bayes factor robustness checks are shown in Figure 3 (theoretical knowledge), Figure 4 (practical knowledge), Figure 5 (critical thinking), Figure 6 (self-awareness), and Figure 7 (citizenship).

XSL•FO
RenderX

Figure 3. Health literacy: theoretical knowledge. BF: Bayes factor.

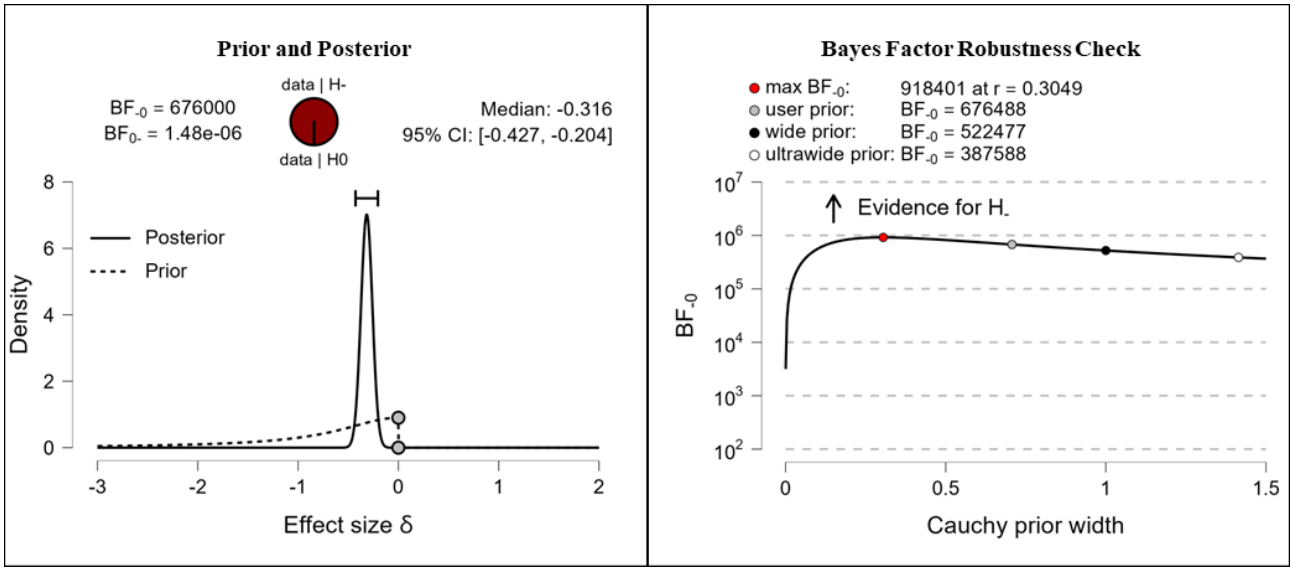


Figure 4. Health literacy: practical knowledge. BF: Bayes factor.

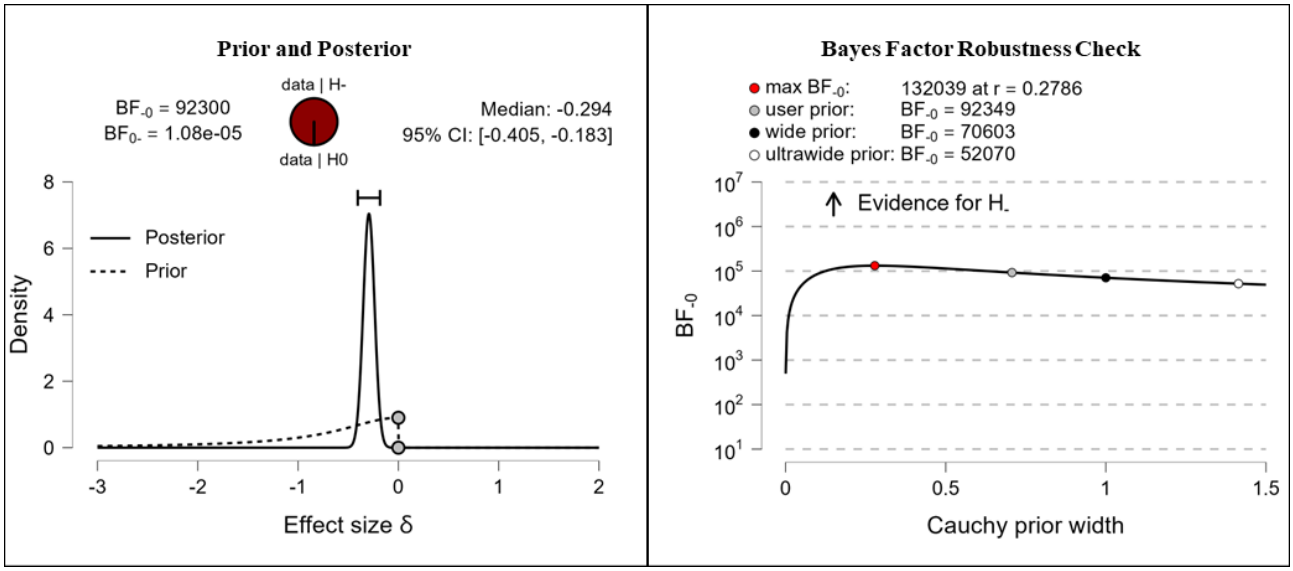
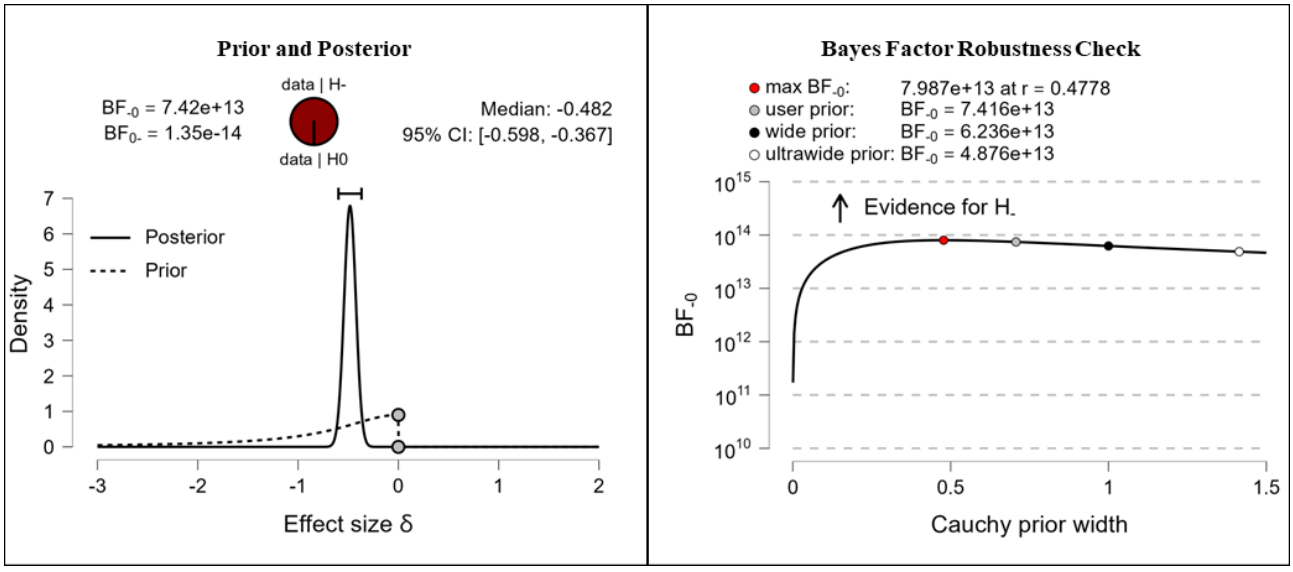


Figure 5. Health literacy: critical thinking. BF: Bayes factor.



XSL•FO
RenderX

Figure 6. Health literacy: self-awareness. BF: Bayes factor.

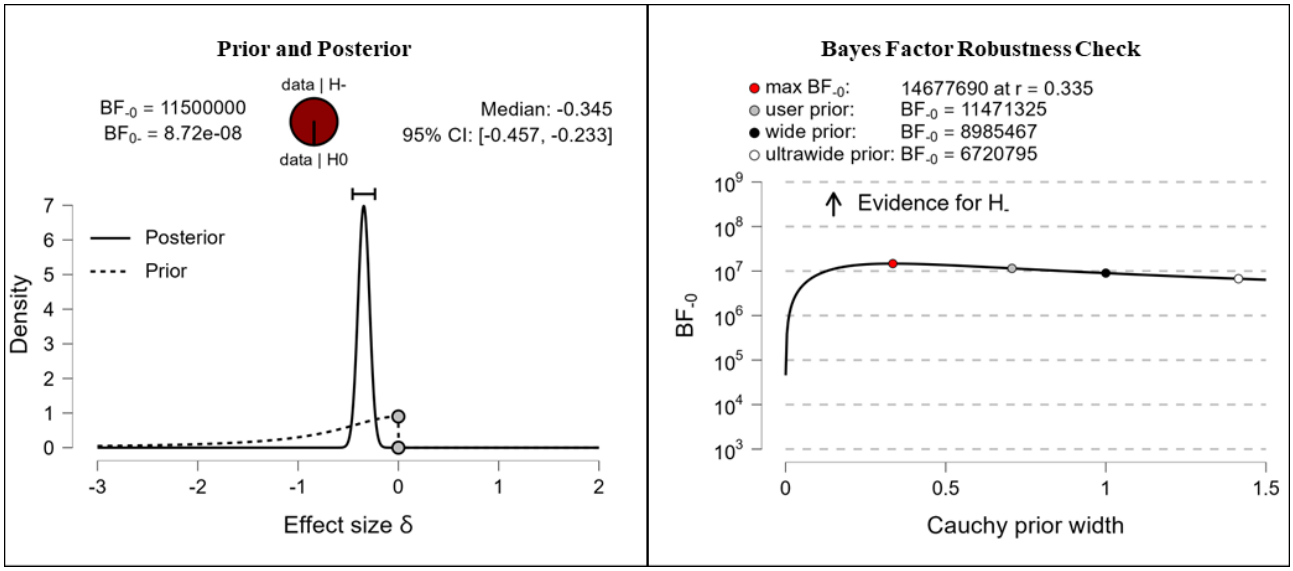
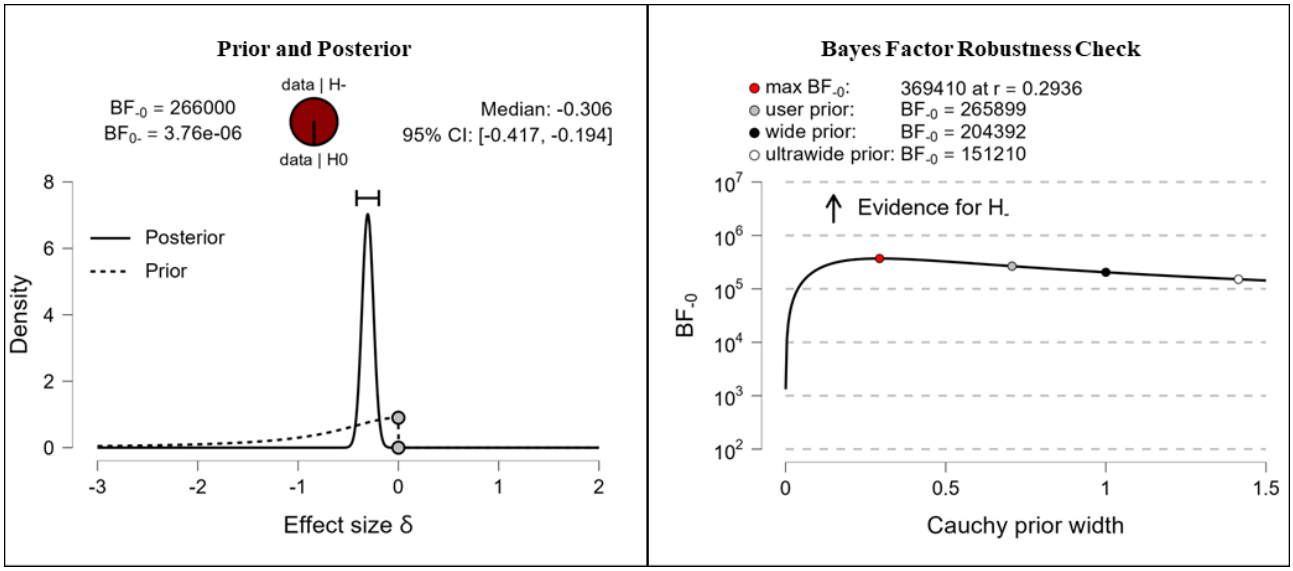


Figure 7. Health literacy: citizenship. BF: Bayes factor.



Digital Health Literacy

It was hypothesized that participants' digital health literacy levels would be higher after completing the e-learning course than they were before completing the e-learning course. Depending on the measured domains, the results vary from anecdotal to extreme evidence for the hypothesis. After completing the e-learning course, participants reported higher competencies in the fields of information searching (anecdotal evidence, BF_{-0}=2.339; δ=–0.135), evaluating reliability (extreme evidence, BF_{-0}=2.03e+11; δ=–0.434), and determining relevance (extreme evidence, BF_{-0}=316,000; δ=–0.308). The corresponding prior and posterior distribution plots, effect sizes, and Bayes factor robustness checks are shown in Figure 8 (information searching), Figure 9 (evaluating reliability), and Figure 10 (determining relevance).

XSL•FO
RenderX

Figure 8. Digital health literacy: information searching. BF: Bayes factor.

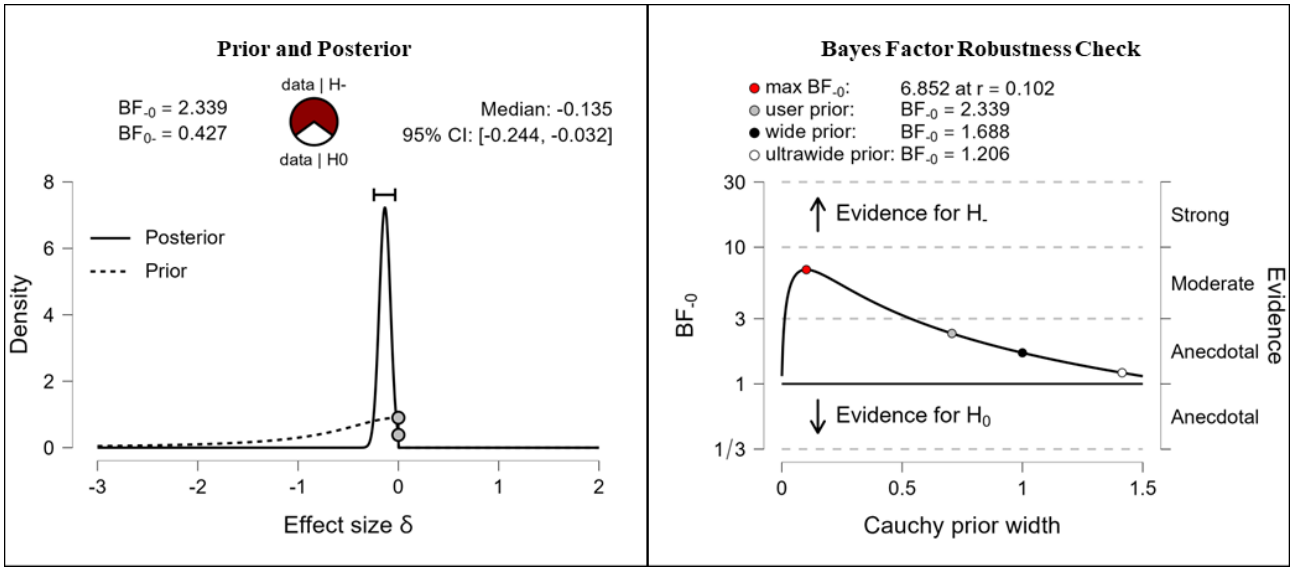


Figure 9. Digital health literacy: evaluating reliability. BF: Bayes factor.

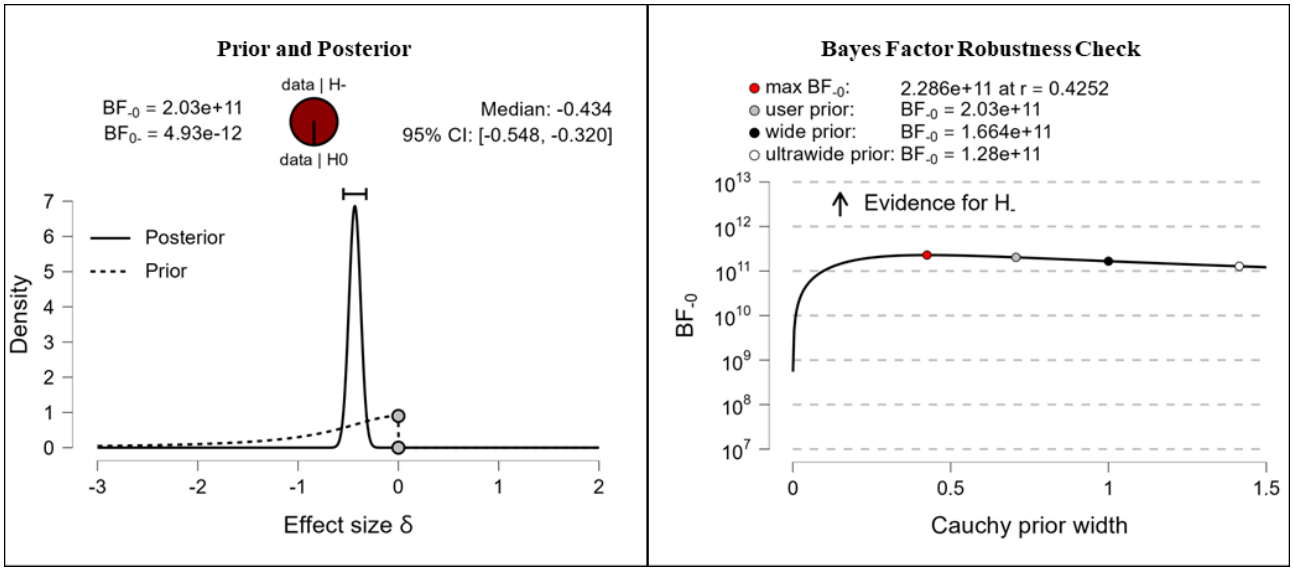
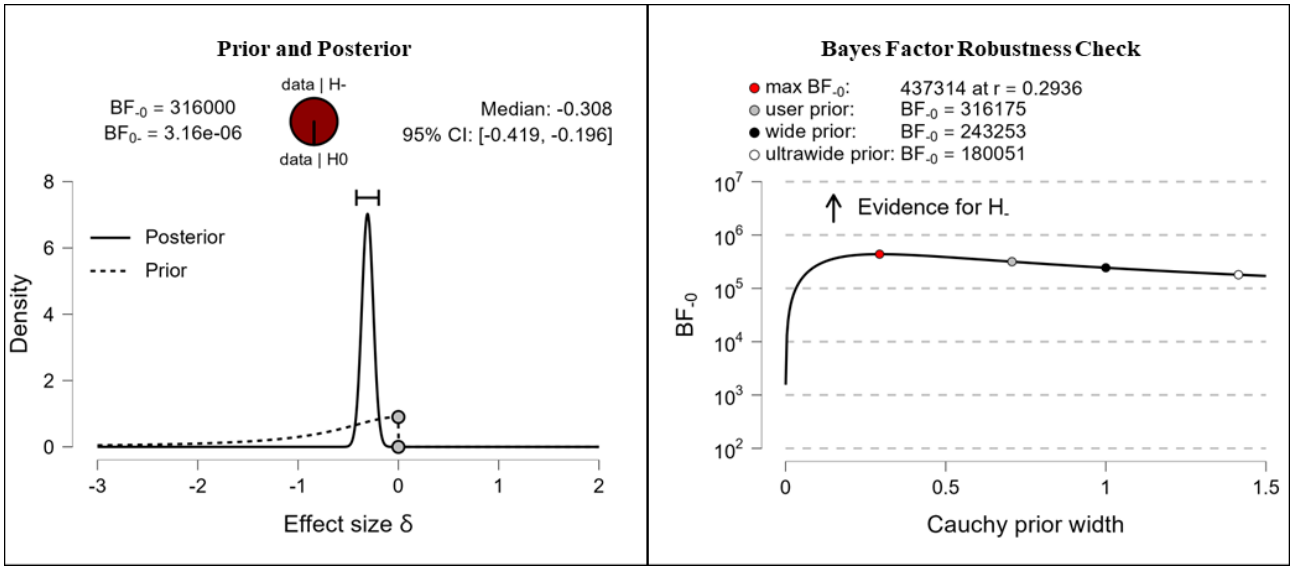


Figure 10. Digital health literacy: determining relevance. BF: Bayes factor.

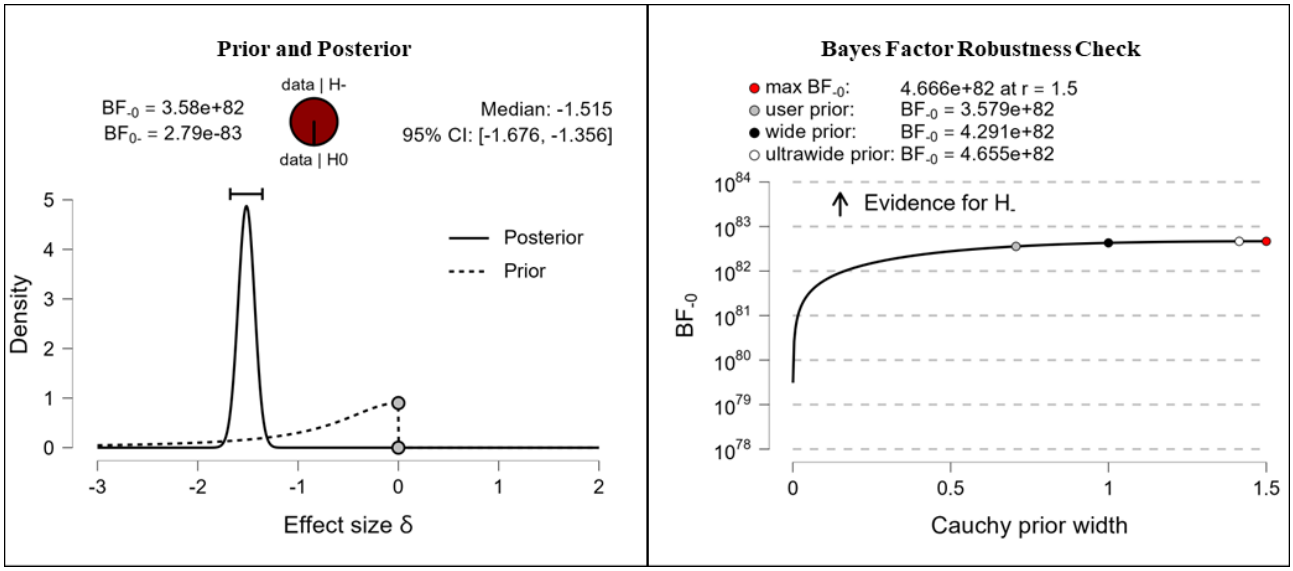


XSL•FO
RenderX

Subjective (Digital) Health Literacy Knowledge

It was hypothesized that participants' subjective knowledge in the domain of (digital) health literacy would be higher after completing the e-learning course than it was before completing the e-learning course. The results show extreme evidence (BF_{-0}=3.58e+82; δ=–1.515) for the hypothesis. The corresponding prior and posterior distribution plot, effect size, and Bayes factor robustness check are shown in Figure 11.

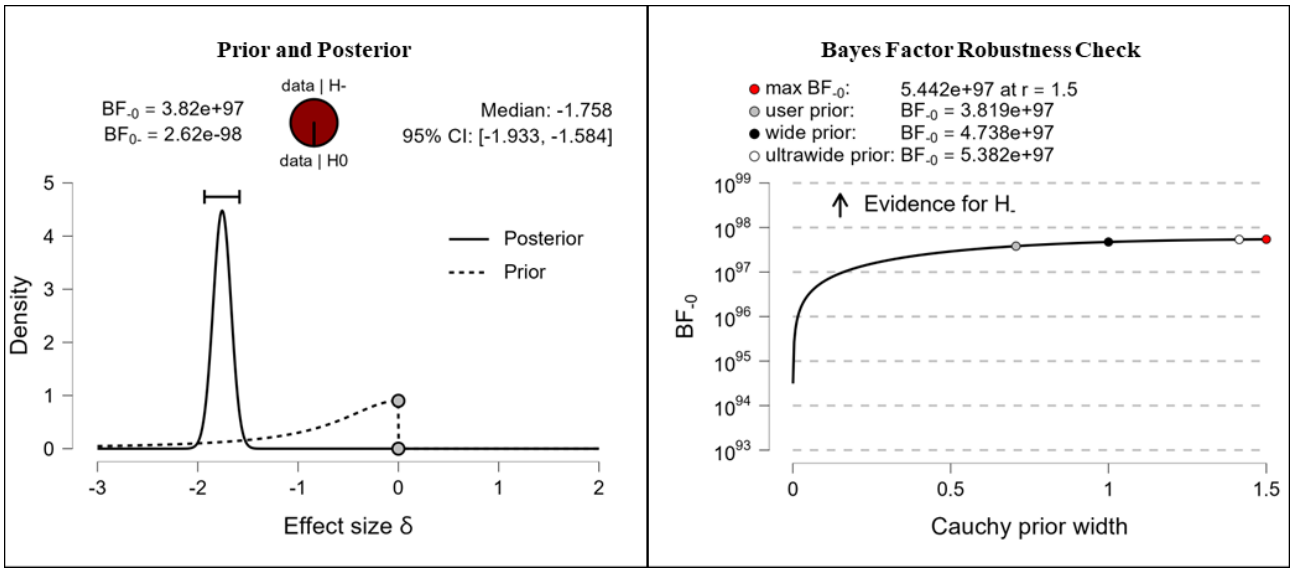
Figure 11. (Digital) health literacy knowledge: subjective. BF: Bayes factor.



Objective (Digital) Health Literacy Knowledge

It was hypothesized that participants' objective knowledge in the domain of (digital) health literacy would be higher after completing the e-learning course than it was before completing the e-learning course. The results show extreme evidence (BF_{-0}=3.82e+97; δ=–1.758) for the hypothesis. The corresponding prior and posterior distribution plot, effect size, and Bayes factor robustness check are shown in Figure 12.

Figure 12. (Digital) health literacy knowledge: objective. BF: Bayes factor.



Discussion

The goal of this study was to evaluate a newly developed e-learning course available on the e-learning platform Gesundweiser.de and its potential to promote (digital) health literacy in school-age children. It was hypothesized that participants' health literacy (hypothesis 1) and digital health literacy levels (hypothesis 2) would be higher after completing the e-learning course than they were before completing the e-learning course. Furthermore, it was hypothesized that participants' subjective (hypothesis 3) and objective knowledge (hypothesis 4) in the domain of (digital) health literacy would be higher after completing the e-learning course than it was before completing the e-learning course. The results support all 4 hypotheses. After completing the e-learning course, participants achieved higher health literacy levels. More specifically, they reached higher competency levels in the domains of theoretical knowledge, practical knowledge, critical thinking, self-awareness, and citizenship. Furthermore, participants achieved higher digital health literacy levels. More

XSL•FO
RenderX

specifically, they reached higher competency levels in the domains of information searching, evaluating reliability, and determining relevance. Moreover, participants demonstrated higher subjective and objective knowledge in the domain of (digital) health literacy after completing the e-learning course.

There are several reasons why these results are encouraging, and 2 of them seem especially important. First, experts have long argued that (digital) health literacy should be taught to school-age children [24-28]. At the same time, however, there are not many German-language interventions available to promote (digital) health literacy that have been scientifically evaluated and proven to work. The e-learning platform Gesundweiser.de closes this gap by providing a scientifically evaluated e-learning course for school-age children that is freely available for pupils, parents, teachers, and all other interested parties. Second, experts have argued that the promotion of (digital) health literacy should be an integral part of the school curriculum [29,30]. However, promoting (digital) health literacy within the German school system is difficult because (digital) health education is not a mandatory school subject in all German states. Therefore, experts suggest that (digital) health literacy could be addressed as part of the mandatory framework for digital education and digital literacy in schools, which was developed by the German Conference on Education Ministries and Cultural Affairs (Kultusministerkonferenz) [32,33]. Because the presented e-learning course was designed to improve (digital) health literacy and at the same time to teach skills specified in the mandatory framework for digital education and digital literacy in schools, it provides an easy way for schools and teachers from all German states to integrate (digital) health literacy education into their school curriculums and lessons.

Even though the results of this study show that the e-learning course available in the e-learning platform Gesundweiser.de has the potential to promote (digital) health literacy in school-age children, there are limitations to the generalizability of the results. Three limitations seem especially important. The first limitation concerns the age of the study participants. All study participants were aged 16 to 20 years. Because previous research has shown that age might influence the suggestibility to misinformation, source monitoring, and digital literacy, the results of this study may not be generalized to younger age groups [45-47]. Therefore, future research should replicate this study with study participants younger than 16 years to explore whether the e-learning course can also promote (digital) health literacy in younger age groups.

The second limitation concerns the methodological approach that was chosen. This study employed a pre-post measurement study design. This means that study participants answered all measures both before and after completing the e-learning course. One of the main advantages of this methodological approach is that it can reduce random noise. In some circumstances, however, this methodological approach might reveal the aim of the study to the participants. After completing questionnaires about their (digital) health literacy and answering knowledge questions about (digital) health literacy, participants might have guessed that the study was designed to test whether the provided e-learning course has the potential to improve (digital) health literacy. This, in turn, might have induced a demand effect that influenced participants' evaluations and learning motivation [48]. Therefore, future studies should test the rationale of this study with a different methodological approach. For example, a between-subject experimental design could be chosen in which participants are randomly assigned to an experimental or control group and answer the dependent measures just once at the end of the study.

The third limitation concerns the setting in which participants could complete the e-learning course. To simulate a homework exercise, participants were given 7 days to complete the course. During this 7-day period, participants could use any device to complete the course and they could start and pause the course as often as they liked. Simulating a homework exercise by giving participants the opportunity to complete the course within 7 days has the advantage that it increases ecological validity. However, it also creates methodological disadvantages. It cannot be guaranteed, for example, that participants completed the course without any help from parents or friends. Furthermore, learning results might be influenced by the number of times participants started and paused the course and by whether participants completed the optional modules. Following the principle of data parsimony, data collection focused on the variables that were most relevant for hypothesis testing and no data were collected regarding the number of times participants started and paused the course and whether participants completed the optional modules. To ensure that participants complete the course without any external help and to investigate the effects of completing the optional modules and pausing and restarting the course, future studies could repeat this study in a laboratory setting and control for the described variables.

Acknowledgments

This research was supported by the independent, nonprofit foundation Stiftung Gesundheitswissen. The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LK, EMB, AE, and RS contributed to the conceptualization of the study and to the study design. SPLENDID RESEARCH GmbH was responsible for data acquisition. LK and EMB were responsible for analysis and interpretation of data. LK was responsible for drafting the manuscript. LK, EMB, AE, and RS were responsible for revising the manuscript critically for important intellectual content. LK, EMB, AE, and RS approved the revision of the manuscript to be published.

XSL•FO
RenderX

Conflicts of Interest

All authors are employees of the independent, nonprofit foundation Stiftung Gesundheitswissen and were involved in the development and evaluation of the e-learning platform Gesundweiser.de.

References

1. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
2. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
3. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
4. Neter E, Brainin E. Association Between Health Literacy, eHealth Literacy, and Health Outcomes Among Patients With Long-Term Conditions. European Psychologist 2019 Jan;24(1):68-81. [doi: 10.1027/1016-9040/a000350]
5. Hurrelmann K, Klinger J, Schaeffer D. Comparison of Health Literacy of the Population in Germany between 2014 and 2020. Gesundheitswesen 2022 Jan 28:1-10. [doi: 10.1055/a-1709-1011] [Medline: 35098501]
6. Schaeffer D, Berens E, Vogt D, Gille S, Griese L, Klinger J, et al. Health Literacy in Germany - Findings of a Representative Follow-up Survey. Dtsch Arztebl Int 2021 Oct 29;118(43):723-728 [FREE Full text] [doi: 10.3238/arztebl.m2021.0310] [Medline: 34551856]
7. Schaeffer D, Berens E, Gille S, Griese L, Klinger J, Sombre SD, et al. Gesundheitskompetenz der Bevölkerung in Deutschland vor und während der Corona Pandemie: Ergebnisse des HLS-GER 2. Bielefeld, Germany: Universität Bielefeld, Interdisziplinäres Zentrum für Gesundheitskompetenzforschung; 2021.
8. Prestin A, Vieux SN, Chou WS. Is Online Health Activity Alive and Well or Flatlining? Findings From 10 Years of the Health Information National Trends Survey. J Health Commun 2015;20(7):790-798. [doi: 10.1080/10810730.2015.1018590] [Medline: 26042588]
9. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. How do people in Germany seek health information? Insights from the first wave of HINTS Germany. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Sep;63(9):1151-1160. [doi: 10.1007/s00103-020-03192-x] [Medline: 32666180]
10. König L, Jucks R. Effects of Positive Language and Profession on Trustworthiness and Credibility in Online Health Advice: Experimental Study. J Med Internet Res 2020 Mar 10;22(3):e16685 [FREE Full text] [doi: 10.2196/16685] [Medline: 32154786]
11. Kolpatzik K, Mohrmann M, Zeeb H. Digitale Gesundheitskompetenz in Deutschland. Berlin, Germany: KomPart; 2020.
12. Samerski S, Müller H. Digital health literacy in Germany - requested, but not supported? Results of the empirical study TK-DiSK. Z Evid Fortbild Qual Gesundhwes 2019 Aug;144-145:42-51. [doi: 10.1016/j.zefq.2019.05.006] [Medline: 31307911]
13. Stadtler M, Winter S, Scharrer L, Thomm E, Krämer N, Bromme R. Selektion, Integration und Evaluation. Psychologische Rundschau 2017 Jul;68(3):177-181. [doi: 10.1026/0033-3042/a000361]
14. König L, Breves P. Providing health information via Twitter: professional background and message style influence source trustworthiness, message credibility and behavioral intentions. JCOM 2021;20(04):A04. [doi: 10.22323/2.20040204]
15. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. Annu Rev Public Health 2020 Apr 02;41:433-451. [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
16. König L, Jucks R. Hot topics in science communication: Aggressive language decreases trustworthiness and credibility in scientific debates. Public Underst Sci 2019 May;28(4):401-416 [FREE Full text] [doi: 10.1177/0963662519833903] [Medline: 30843467]
17. König L, Jucks R. Influence of Enthusiastic Language on the Credibility of Health Information and the Trustworthiness of Science Communicators: Insights From a Between-Subject Web-Based Experiment. Interact J Med Res 2019 Aug 12;8(3):e13619 [FREE Full text] [doi: 10.2196/13619] [Medline: 31411138]
18. König L, Jucks R. When do information seekers trust scientific information? Insights from recipients' evaluations of online video lectures. Int J Educ Technol High Educ 2019 Feb 13;16(1):1-21. [doi: 10.1186/s41239-019-0132-7]
19. König L. Podcasts in higher education: teacher enthusiasm increases students' excitement, interest, enjoyment, and learning motivation. Educational Studies 2020 Jan 07;47(5):627-630. [doi: 10.1080/03055698.2019.1706040]
20. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
21. Dadaczynski K, Okan O, Messer M, Leung AYM, Rosário R, Darlington E, et al. Digital Health Literacy and Web-Based Information-Seeking Behaviors of University Students in Germany During the COVID-19 Pandemic: Cross-sectional Survey Study. J Med Internet Res 2021 Jan 15;23(1):e24097 [FREE Full text] [doi: 10.2196/24097] [Medline: 33395396]

XSL•FO
RenderX

22. König L. Lehramtsstudierende: Ihre Meinungen zur Digitalisierung im Bildungssystem und zu den Unterrichtsfächern Programmieren und Unternehmertum. MedienPädagogik 2020 May 11:68-85. [doi: 10.21240/mpaed/00/2020.05.11.x]
23. OECD. 21st-Century Readers: Developing Literacy Skills in a Digital World. Paris, France: OECD Publishing; 2021.
24. Bröder J, Okan O, Bollweg TM, Bruland D, Pinheiro P, Bauer U. Child and Youth Health Literacy: A Conceptual Analysis and Proposed Target-Group-Centred Definition. Int J Environ Res Public Health 2019 Sep 14;16(18):3417 [FREE Full text] [doi: 10.3390/ijerph16183417] [Medline: 31540040]
25. Okan O. The importance of early childhood to address equity and health literacy development in the life-course. Public Health Panorama 2019;5(2):170-176 [FREE Full text]
26. Paakkari L, Inchley J, Schulz A, Weber M, Okan O. Addressing health literacy in schools in the WHO European Region. Public Health Panorama 2019;5:186-190 [FREE Full text]
27. Videto DM, Dake JA. Promoting Health Literacy Through Defining and Measuring Quality School Health Education. Health Promot Pract 2019 Nov;20(6):824-833 [FREE Full text] [doi: 10.1177/1524839919870194] [Medline: 31465242]
28. St Leger L. Schools, health literacy and public health: possibilities and challenges. Health Promot Int 2001 Jun;16(2):197-205. [doi: 10.1093/heapro/16.2.197] [Medline: 11356758]
29. World Health Organization. Shanghai declaration on promoting health in the 2030 Agenda for Sustainable Development. Health Promot Int 2017 Feb 01;32(1):7-8. [doi: 10.1093/heapro/daw103] [Medline: 28180270]
30. Health literacy in the context of health, well-being and learning outcomes – the case of children and adolescents in schools. World Health Organization Regional Office for Europe. 2021. URL: https://tinyurl.com/3mv8cns9 [accessed 2022-05-11]
31. Paakkari L, Torppa M, Mazur J, Boberova Z, Sudeck G, Kalman M, et al. A Comparative Study on Adolescents' Health Literacy in Europe: Findings from the HBSC Study. Int J Environ Res Public Health 2020 May 19;17(10):3543 [FREE Full text] [doi: 10.3390/ijerph17103543] [Medline: 32438595]
32. Kultusministerkonferenz. Bildung in der digitalen Welt. Berlin, Germany: Sekretariat der Kultusministerkonferenz; 2016.
33. Schulenkorf T, Krah V, Dadaczynski K, Okan O. Addressing Health Literacy in Schools in Germany: Concept Analysis of the Mandatory Digital and Media Literacy School Curriculum. Front Public Health 2021;9:687389 [FREE Full text] [doi: 10.3389/fpubh.2021.687389] [Medline: 34291029]
34. Paakkari L, Okan O. Health Literacy-Talking the Language of (School) Education. Health Lit Res Pract 2019 Jul;3(3):e161-e164 [FREE Full text] [doi: 10.3928/24748307-20190502-01] [Medline: 31410386]
35. Herlitz L, MacIntyre H, Osborn T, Bonell C. The sustainability of public health interventions in schools: a systematic review. Implement Sci 2020 Jan 06;15(1):4 [FREE Full text] [doi: 10.1186/s13012-019-0961-8] [Medline: 31906983]
36. Dadaczynski K, Paulus P, Nieskens B, Hundeloh H. Gesundheit im Kontext von Bildung und Erziehung – Entwicklung, Umsetzung und Herausforderungen der schulischen Gesundheitsförderung in Deutschland. Z f Bildungsforsch 2015 Jan 28;5(2):197-218. [doi: 10.1007/s35834-015-0122-3]
37. Lühnen J, Albrecht M, Hanßen K, Hildebrandt J, Steckelberg A. Guideline for the Development of Evidence-based Patient Information: insights into the methods and implementation of evidence-based health information. Z Evid Fortbild Qual Gesundhwes 2015;109(2):159-165. [doi: 10.1016/j.zefq.2015.03.004] [Medline: 26028454]
38. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
39. Paakkari O, Torppa M, Kannas L, Paakkari L. Subjective health literacy: Development of a brief instrument for school-aged children. Scand J Public Health 2016 Dec;44(8):751-757. [doi: 10.1177/1403494816669639] [Medline: 27655781]
40. Felder-Puig R, Teutsch F, Ramelow D, Maier G. Gesundheit und Gesundheitsverhalten von österreichischen Schülerinnen und Schülern. Vienna, Austria: Bundesministerium für Arbeit, Soziales, Gesundheit und Konsumentenschutz; 2019.
41. JASP team. JASP, version 0.16.1 (computer software). Amsterdam, The Netherlands: University of Amsterdam; 2022.
42. Kelter R. Bayesian alternatives to null hypothesis significance testing in biomedical research: a non-technical introduction to Bayesian inference with JASP. BMC Med Res Methodol 2020 Jun 05;20(1):142 [FREE Full text] [doi: 10.1186/s12874-020-00980-6] [Medline: 32503439]
43. van Doorn J, van den Bergh D, Böhm U, Dablander F, Derks K, Draws T, et al. The JASP guidelines for conducting and reporting a Bayesian analysis. Psychon Bull Rev 2021 Jun;28(3):813-826 [FREE Full text] [doi: 10.3758/s13423-020-01798-5] [Medline: 33037582]
44. Quintana DS, Williams DR. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. BMC Psychiatry 2018 Jun 07;18(1):178 [FREE Full text] [doi: 10.1186/s12888-018-1761-4] [Medline: 29879931]
45. Choi W, Stvilia B. Web credibility assessment: Conceptualization, operationalization, variability, and models. J Assn Inf Sci Tec 2015 May 13;66(12):2399-2414. [doi: 10.1002/asi.23543]
46. Jin K, Reichert F, Cagasan LP, de la Torre J, Law N. Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. Computers & Education 2020 Nov;157:103968. [doi: 10.1016/j.compedu.2020.103968]
47. Mitchell KJ, Johnson MK, Mather M. Source monitoring and suggestibility to misinformation: adult age-related differences. Appl Cognit Psychol 2002 Jan;17(1):107-119. [doi: 10.1002/acp.857]

XSL•FO
RenderX

48. Charness G, Gneezy U, Kuhn MA. Experimental methods: Between-subject and within-subject design. J Econ Behav Organ 2012 Jan;81(1):1-8. [doi: 10.1016/j.jebo.2011.08.009]

Abbreviations

BF: Bayes factor

Edited by A Mavragani; submitted 23.02.22; peer-reviewed by T Schulenkorf, K Cronan, CF Lai; comments to author 18.03.22; accepted 19.04.22; published 16.05.22.

Please cite as:
König L, Marbach-Breitrück E, Engler A, Suhr R
The Development and Evaluation of an e-Learning Course That Promotes Digital Health Literacy in School-age Children: Pre-Post Measurement Study
J Med Internet Res 2022;24(5):e37523
URL: https://www.jmir.org/2022/5/e37523
doi:10.2196/37523
PMID:35576572

©Lars König, Eugenia Marbach-Breitrück, Anne Engler, Ralf Suhr. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.05.2022. 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.

XSL•FO
RenderX

Original Paper

A Short Intervention and an Interactive e-Learning Module to Motivate Medical and Dental Students to Enlist as First Responders: Implementation Study

Victor Taramarcaz[1*], BSc; Tara Herren[1*], MSc; Eric Golay[1], DAS; Simon Regard[1], MD; Sébastien Martin-Achard[2], DAS; Francois Mach[3], MD; Nicolas Schnetzler[1], BSc; Gaëtan Ricci[1], BSc; Ido Zamberg[4], MD; Robert Larribau[1], MD; Marc Niquille[1], MD; Mélanie Suppan[4], MD; Eduardo Schiffer[4], MD, PhD; Laurent Suppan[1], MD

[1]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Save a Life, Swiss Emergency Responders Association, Geneva, Switzerland

[3]Cardiology Department, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[4]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
University of Geneva Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 223723311
Email: laurent.suppan@hcuge.ch

Abstract

Background: Prompt and proficient basic life support (BLS) maneuvers are essential to increasing the odds of survival after out-of-hospital cardiac arrest. However, significant time can elapse before the arrival of professional rescuers. To decrease these delays, many countries have developed first responder networks. These networks are composed of BLS-certified lay or professional rescuers who can be dispatched by emergency medical communication centers to take care of those who experience out-of-hospital cardiac arrest. Many systems are, however, limited by a relatively low number of active first responders, and first-year medical and dental students may represent an almost untapped pool of potential rescuers. On top of providing an enhanced BLS coverage to the population, this could also help medical students be better prepared to their future role as certified health care providers and address societal expectations regarding health care students.

Objective: Our objective was to describe the impact of a short motivational intervention followed by a blended BLS course (e-learning and practice session) designed to motivate first-year medical and dental students to enlist as first responders.

Methods: A short, web-based, motivational intervention presenting this project took place, and first-year University of Geneva, Faculty of Medicine students were provided with a link to the study platform. Those who agreed to participate were redirected to a demographic questionnaire before registering on the platform. The participants were then asked to answer a second questionnaire designed to determine their baseline knowledge prior to following an interactive e-learning module. Upon completion, a web-based booking form enabling them to register for a 1-hour practice session was displayed. These sessions were held by senior medical students who had been trained and certified as BLS instructors. The participants who attended these practice sessions were asked to answer a postcourse questionnaire before receiving the certificate enabling them to register as first responders.

Results: Out of the 529 first-year students registered at University of Geneva, Faculty of Medicine on January 14, 2021, 190 (35.9%) initially agreed to participate. Moreover, 102 (19.3%) attended the practice sessions, and 48 (9.1%) had completed all training and enlisted as first responders on the dedicated platform, Save a Life, at 6 months (July 14, 2021). Postcourse confidence in resuscitation skills was associated with a higher likelihood of registering as first responder (P=.03). No association was found

XSL•FO
RenderX

between prior BLS knowledge and the probability of registering to a practice session (*P*=.59), of obtaining a course completion certificate (*P*=.29), or of enlisting as first responder (*P*=.56).

Conclusions: This study shows that a motivational intervention associated with a short BLS course can convince medical students to enlist as first responders. Further studies are needed to understand the rather low proportion of medical students finally registering as first responders.

International Registered Report Identifier (IRRID): RR2-10.2196/24664

(J Med Internet Res 2022;24(5):e38508) doi:10.2196/38508

KEYWORDS

basic life support; cardiopulmonary resuscitation; first responder; undergraduate medical education; out-of-hospital cardiac arrest; medical education; e-learning; digital education; medical student; blended learning

Introduction

Background

According to the World Health Organization, ischemic heart disease is the current leading cause of death worldwide [1], and most out-of-hospital cardiac arrests (OHCAs) are linked to this condition [2]. Survival and neurological outcomes greatly depend upon the prompt provision of basic life support (BLS) maneuvers and on the availability of automatic external defibrillators (AED) [3-6]. In Geneva, Switzerland, BLS maneuvers were only provided in 40% of OHCA cases between 2009 and 2012, according to a retrospective study published in 2018 [7]. This proportion is lower than that found on average in Europe [2], and there is, consequently, considerable room for improvement.

To enable the provision of BLS maneuvers prior to the arrival of professional rescuers, first responder systems have been created in many regions of the world. These systems allow emergency medical dispatchers to send adequately trained individuals to take care of OHCA situations before the arrival of professional help. Although different technological solutions have been developed, some of which are currently used in Switzerland, most operate according to the same principles. In general, first responders receive a notification on their smartphone through a specific app. The ones who are both available and within a 3-km radius of the scene receive the exact coordinates after accepting the mission. This allows a timely provision of BLS maneuvers by adequately trained responders and, consequently, higher survival rates [8-12]. The major functional limitation of these platforms is related to the number of registered first responders, their location, and their availability. By raising the sheer number of first responders, a country or state could increase its overall coverage and henceforth improve OHCA outcomes. In Switzerland, the state of Ticino was first to launch a first responder system in 2005 [13]. Since then, other systems have been developed throughout the country, and most of the 26 Swiss states now run their own first responder organization [14]. In Geneva, the first responder platform, *Save a Life*, enables any adult in possession of a valid BLS-AED certificate to register as first responder [15]. Recently created, *Save a Life* counted only 260 active first responders in 2019, its first year of activity [16].

Medical students may represent an almost untapped pool of potential first responders. Previous studies have shown that first-year medical students feel that they would be expected to act in case of an emergency but feel unprepared to face OHCA situations [17]. This lack of confidence seems legitimate as their BLS knowledge is not much more advanced than that of the general population [4]. Giving these novice students the opportunity to gain solid knowledge about first aid procedures could increase their confidence, motivating them to join a first responder platform and thereby potentially increasing the rate of BLS maneuvers initiated prior to the arrival of professional rescuers.

Objective

The main objective of this study was to describe the impact of implementing a short motivational intervention and a blended learning path designed to motivate first-year medical and dental students to enlist as first responders. To be able to enlist as first responders, students were required to follow an e-learning module, attend a practice session, and answer multiple questionnaires.

Methods

Study Design

This was a prospective implementation study based on a research protocol published on November 6, 2020 [18]. When relevant, methods and results are reported according to the Checklist for Reporting Results of Internet E-Surveys [19].

Ethical Considerations

A declaration of “no objection” was issued by the local ethics committee (Req-2020-01143) as this project did not fall within the scope of the Swiss Federal Act on Research involving Human Beings [20]. This project was also approved by the vice dean of undergraduate education at the University of Geneva, Faculty of Medicine (UGFM).

Participants and Enrollment

The target population consisted in a convenience sample including all UGFM first-year medical and dental students. Those who were already registered as first responders were excluded. No financial incentive was given to promote participation. The whole learning path was completely free, and students were informed that they would be granted a specific BLS-AED certificate upon a successful completion of the training program. This certificate had a 1-year validity and only allowed them to register on the *Save a Life* platform.

XSL•FO
RenderX

Two senior medical students presented the project to all these students on January 14, 2021. The presentation, which was originally intended to take place live in an auditorium at the beginning of a lecture about atherosclerosis, was held online because of restrictions linked to the COVID-19 pandemic. It was broadcasted as part of this lecture through the university's web-based platform on which all courses could be followed live or on-demand. The presentation included an overview of the project and the learning path, an estimate of its duration, and a presentation of the *Save a Life* first responder system [15]. The URL linking to the study site was also shown on the last slide of the presentation. The study URL was also sent to all medical and dental students through a mailing list a few hours after the presentation (Multimedia Appendix 1). Theoretically, this official UGFM mailing list should have included all first-year medical and dental students. However, subscribing to this list was not mandatory, and the university's policy allows anyone in possession of a University of Geneva email address to subscribe to any mailing list. This made it impossible to ascertain that the whole population could be reached. A second and last email reminder promoting registration was sent to the whole mailing list on March 8, 2021.

Web-Based Platform

A specific web-based study platform was developed using the Joomla 3.9 [21] content management system (Open Source Matters). This platform hosted the questionnaires, the e-learning module, and was used to manage the registration process for the practice sessions. All questionnaires were created using Community Surveys 5.5 (CoreJoomla). Three authors thoroughly tested the platform prior to study inception. All collected data were stored on an encrypted MySQL database (MariaDB 5.5.5) located on a Swiss server and handled in accordance with the European General Data Protection Regulation [22].

Consent and Initial Questionnaire

The URL provided to the students led to an introductory page designed to determine whether they were already registered as first responders, which was the first of the two exclusion criteria (Table 1 and Multimedia Appendix 2). Those who met this first criterion were nevertheless given the possibility of following the e-learning module. All the other participants were redirected to a form designed to gather their informed consent and determine whether our second and last exclusion criteria were met (ie, not being part of the target population). A disclaimer (Multimedia Appendix 3) detailing the study design and data handling procedures was displayed at the top of the main page, and participants had the possibility of reading and downloading a comprehensive 3-page document (Multimedia Appendix 4) including additional details. The contact information of 4 investigators was provided to enable them to directly ask further questions.

Students who refused to participate were prompted to give a reason and were nevertheless allowed to access the e-learning module (Table 1 and Multimedia Appendix 2). They were offered the possibility of following the practice sessions upon completion of the module and could therefore obtain a course completion certificate and join the first responder system regardless of their willingness to participate in the study.

Those who did not meet the exclusion criteria and agreed to participate were redirected to a short registration form, which was composed of 3 fields only: first name, last name, and email address. The students' identities were collected to allow the creation of nominative certificates. Email addresses were used to directly contact the participants, give them information regarding the practice sessions, and send them their BLS certificate provided they had successfully completed the whole learning path. The participants were informed that they could withdraw from the study at any time.

After registration, the participants were asked to fill out a precourse questionnaire (Table 2 and Multimedia Appendix 5) designed to gather demographic data and assess initial BLS-AED knowledge. The questionnaire was adapted from a study by Sturny et al [4] and based on the 2015 International Consensus on Cardiopulmonary Resuscitation and Emergency Cardiovascular Care Science With Treatment Recommendations [6].

Table 1. Screening questionnaire and consent form (survey page number 1).

Survey field and question	Type of question
Already filled the questionnaire or exclusion criteria	
Already a first responder?	Yes/no
Demographics	
Student at UGFM[a]?	Yes/no
If no: current professional status?	Open
Consent	
Agree to participate?	Yes/no
If no: reasons for refusal?	MAQ[b]
If no: access to the e-learning module?	Yes/no

[a]UGFM: University of Geneva, Faculty of Medicine.

[b]MAQ: multiple answer question.

XSL•FO
RenderX

Table 2. Precourse questionnaire.

Survey page, field, and question	Type of question
1—Demographics	
Year of birth	Open (Regex[a])
Gender	MCQ[b]
Medical, biomedical, or dental medicine student	MCQ
Former student or graduate of another health care profession	MCQ
Target specialty	MCQ
2—General BLS[c] knowledge	
Ever heard of BLS or ACLS[d] before	Yes/no
Meaning of AED[e,f]	Open
Year of the last BLS guidelines update	Open (Regex)
Phone number of the emergency medical communication center[f,g]	Open
3—Prior BLS experience	
Prior BLS training	MAQ[h]
Wish for additional BLS training	Yes/no
4—Specific BLS knowledge	
Criteria used to recognize OHCA[f,g,i]	MAQ
BLS-sequence[f,g]	Ordering
Artery for pulse assessment[f]	MCQ
Compression depth[f,g]	MCQ
Compressions: ventilation ratio[f]	MCQ
Compression rate[f,g]	MCQ
Compression-only CPR[f,g,j]	Yes/no
Foreign body airway obstruction[f]	MCQ

[a]A Regex validation rule was used to avoid invalid entries.
[b]MCQ: multiple choice question (only one answer accepted).
[c]BLS: basic life support.
[d]ACLS: advanced cardiovascular life support.
[e]AED: automatic external defibrillator.
[f]Items used to calculate the 10-point score (initial BLS knowledge).
[g]Items used to calculate the 6-point score (essential BLS knowledge).
[h]MAQ: multiple answer question (more than one answer accepted).
[i]OHCA: out-of-hospital cardiac arrest.
[j]CPR: cardiopulmonary resuscitation.

e-Learning and Practice Sessions

After completing this questionnaire, participants were granted access to an interactive e-learning module developed under Storyline 3 (Articulate Global, LLC). This e-learning was adapted from a similar module used to teach BLS-AED procedures to second-year UGFM students. It was designed to last 30 minutes, but no time limit was set for its completion.

The objectives of this e-learning module were designed according to the Swiss Resuscitation Council guidelines for the training of BLS-AED providers [23] and adapted according to the COVID-19 pandemic guidelines [24]. The goal of this module was to enable students to (1) identify a cardiorespiratory arrest, (2) recall the emergency numbers and alert professional help, (3) identify threats to their own safety, (4) acquire the knowledge necessary to perform high-quality cardiopulmonary resuscitation (CPR), and (5) use an AED. To ensure an optimal focus, all these subjects were presented in an interactive way.

After completing the e-learning module, the participants were able to register for the practice sessions. These sessions lasted

XSL•FO
RenderX

1 hour and were limited to 4 participants according to the regional COVID-19 regulations in effect at the time [25,26]. Seventeen senior medical students acted as instructors during these sessions. For the purpose of this study, these senior students had been trained and certified as BLS-AED instructors according to the Swiss Resuscitation Council guidelines between January and March 2021. Owing to the availability of training rooms and instructors, a total of 128 training slots were available between February 8, 2021, and April 30, 2021. The instructors were provided with a comprehensive checklist to ensure that all critical points had been covered and mastered by the participants even though the actual structure of the session was left at their discretion.

Final Questionnaire and Certification

The participants who successfully completed the practice sessions were sent an email inviting them to fill a postcourse questionnaire (Table 3 and Multimedia Appendix 6). A course completion certificate was delivered to all the participants who completed this last questionnaire.

Table 3. Postcourse questionnaire.

Survey page, field, and question	Type of question
1—Opinion	
Appreciation	Yes/no
If yes: positive thoughts	MAQ[a]
If no: negative thoughts	MAQ
General comments	Free text
2—Confidence	
Precourse confidence for OHCA[b] management	Likert scale (1-5)
Postcourse confidence for OHCA management	Likert scale (1-5)
Factors contributing to confidence	Likert scale (1-5)
Factors contributing to lack of confidence	Likert scale (1-5)
Other comments on confidence	Free text
3—First responders	
Intention to register as first responder	Yes/no
If yes: contributing factors	Likert scale (1-5)
If no: impeding factors	Likert scale (1-5)
Other factors	Free text
4—Improvement	
Suggestion for improvement	Free text

[a]MAQ: multiple answer question.

[b]OHCA: out of hospital cardiac arrest.

Deviations From the Research Protocol

The research protocol had to be adapted to cope with unforeseen realities. First, since the course during which the study was presented was web-based, a QR code was not created, and a short URL was displayed. Second, Joomla 3.10 had not been released at the time this study started, and the platform was developed under Joomla 3.9. In addition, we added a postcourse questionnaire after the publication of the research protocol to evaluate the participants' confidence and gather information regarding the practice sessions. Therefore, minor modifications were made to the second questionnaire for consistency purposes. Furthermore, in line with the previous study by Sturny et al [4], we computed a 6-point score focusing on "essential BLS-knowledge." Finally, rather than using already-trained BLS instructors, we decided to specifically train senior medical students for that purpose.

Outcomes

The primary outcome was the number of first-year students who had enlisted as first responders on the *Save a Life* platform by July 14, 2021 (ie, exactly 6 months after the course presenting our study was held). Secondary outcomes were the number of participants who agreed to participate, registered on the platform, began the e-learning module, completed this module, registered for practice sessions, attended these sessions, and obtained a certificate. The evolution of their confidence and the association of postcourse confidence with the probability of registering on the *Save a Life* platform were also analyzed.

Even though the web-based platform was thoroughly tested before study inception, we could not rule out the occurrence of technical difficulties. Therefore, all technical difficulties were recorded and reviewed. Free comments were also analyzed.

XSL•FO
RenderX

Statistical Analysis

Data were extracted to comma-separated value files, and Stata 17.0 (StataCorp LLC) was used for data curation and statistical analysis. The curated data file is available as Multimedia Appendix 7. The different proportions of students are presented using descriptive statistics (n [%]). The items used to calculate the prior BLS knowledge scores can be found in Table 2. Each item was worth 1 point, with no differential weighting. After assessing the normality (first graphically, then using the Kolmogorov-Smirnov test), Student *t* test was used to determine whether there was a difference in the interest in further training according to prior BLS knowledge (using both the 10-point and 6-point scores). The same test was used to assess the probability of obtaining a certificate and of registering as first responder according to prior BLS knowledge. The Student *t* test was also used to describe whether postcourse confidence was different between those who enlisted as first responders and those who did not. A paired *t* test was performed to assess the evolution of student confidence before and after the course. Factors reported as affecting confidence in resuscitation skills and motivation to enlist as first responders according to the questions based on the 5-point Likert scales are described graphically. The preplanned sensitivity analysis—which was designed to determine if differences existed between the students who immediately accepted to participate and those who rallied the study after following the e-learning module—was not performed as there were no participants in this latter group. A *P* value of less than .05 was considered significant.

In addition to the preplanned analyses, we performed a qualitative analysis of the comments obtained through the postcourse questionnaire. The participants were able to leave free comments in 4 different sections (Table 3). Because responses were sometimes entered in inappropriate sections, all comments were pooled before being sorted by theme. Representative answers were translated using DeepL Translator (DeepL) and added to the results section. We also analyzed the emails sent by medical students regarding the technical difficulties they had encountered.

Results

A total of 529 medical and dental students were registered at UGFM at the time this study started (8 more than that stated in the original study protocol). Figure 1 shows the different steps and the associated proportions. By July 14, 2021, 48/529 (9.1%) students had completed the whole process and enlisted as first responders on the *Save a Life* platform. This was a slightly lower proportion than initially expected according to the original protocol (10%, 53/529) [18].

The characteristics of the 162/529 (30.6%) students who completed the first questionnaire are detailed in Table 4. The respondents unanimously answered that they wished for more BLS training. No association was found between prior BLS knowledge and the probability of registering to a practice session (*P*=.59), of obtaining a course completion certificate (*P*=.29), or of enlisting as first responder on the *Save a Life* platform (*P*=.56). The proportion of dental students who completed the first questionnaire was significantly lower than that of medical students (7/55, 12.7% vs 155/474, 32.7%; *P*=.002).

Of the 124 participants who completed the e-learning module, 28 (22.6%) reported problems with the component used to register for practice sessions. Help was provided by email, and all problems were solved manually by the study team. Moreover, 10/107 (9.3%) students who had registered for a practice session did not show up. They were all contacted to determine the reason preventing them from attending the practice sessions, but none of them answered our request for information.

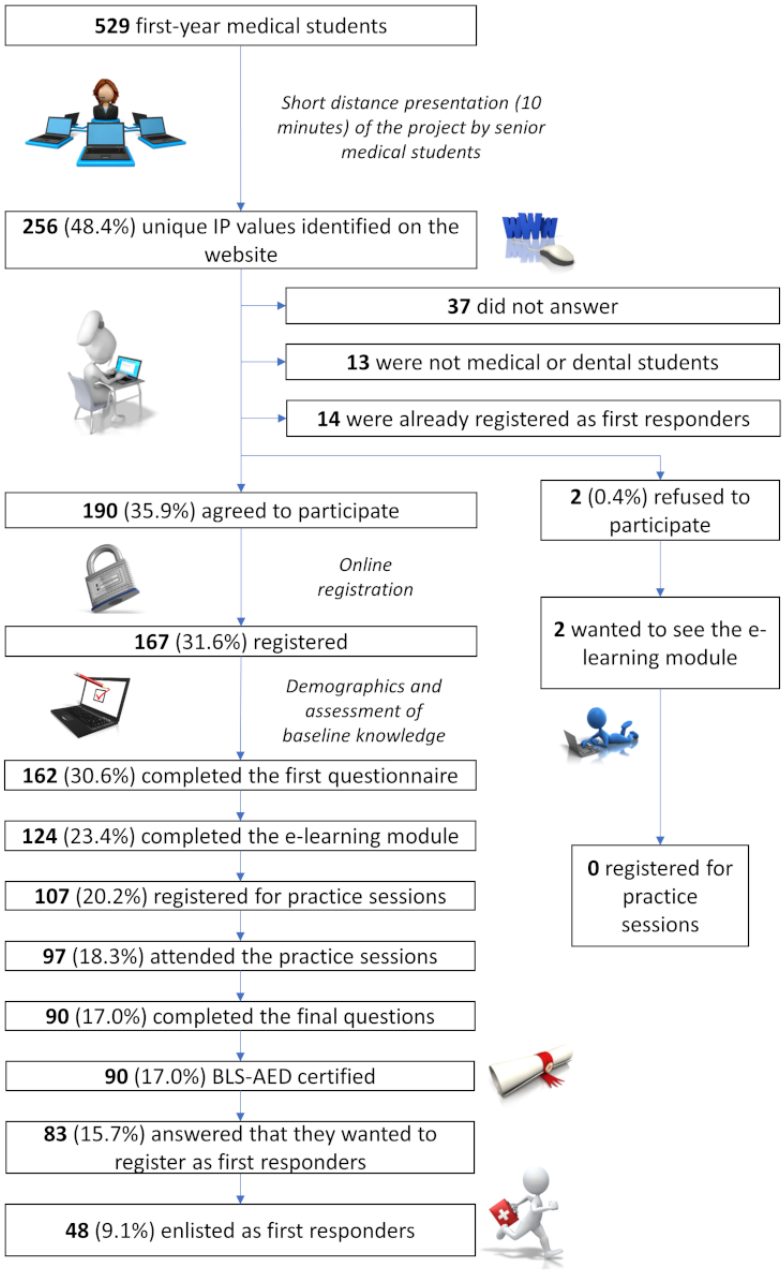
Most of those who attended the practice sessions answered our postcourse questionnaire (90/97, 92.8%). There was a significant increase in confidence regarding OHCA management skills after following the learning path (4.2, SD 0.6 vs 2.1, SD 0.9; *P*<.001). The majority of the students who answered the postcourse questionnaire reported that they were willing to sign up as first responders (83/90, 92.2%). Factors affecting student confidence and their motivation to enlist as first responders are shown in Figures 2-5.

Those who felt more confident after the course were more likely to register as first responders (registrants had a mean confidence of 4.3, SD 0.5 vs 4.0, SD 0.6; *P*=.03).

All the students who completed the postcourse questionnaire had a positive opinion of the learning path (90/90, 100%) and all of them would have recommended this course to other first-year students. Moreover, 30/90 (33.3%) students left a total of 40 comments in the dedicated sections of the postcourse questionnaire. Most were positive feedbacks regarding the project (29/40, 72.5%). Three other themes were identified, which were course duration, integration of the course in the standard curriculum, and availability of a face-to-face course during the COVID-19 pandemic. Many students (15/40, 37.5%) thought that the course was too short and that there was not enough time for hands-on practice. One student commented, “1 hour is too short! Sessions scheduled for a little longer (1h30-2h) would allow us to be less stressed by time and to practice better.“ Some students (5/40, 12.5%) considered that such a course should be mandatory during the first year of their curriculum. Finally, a few students (3/40, 7.5 %) declared that attending this course had been a unique occasion to have face-to-face training during the COVID-19 pandemic. A student commented that “It was really super interesting and rewarding!!! And it was really nice to be able to do some practice,” while another wrote that it was “A pleasure to see people in these times.”

XSL•FO
RenderX

Figure 1. Study flowchart. AED: automatic external defibrillator; BLS: basic life support.



XSL•FO
RenderX

Table 4. Characteristics of the 162 participants who answered the first questionnaire.

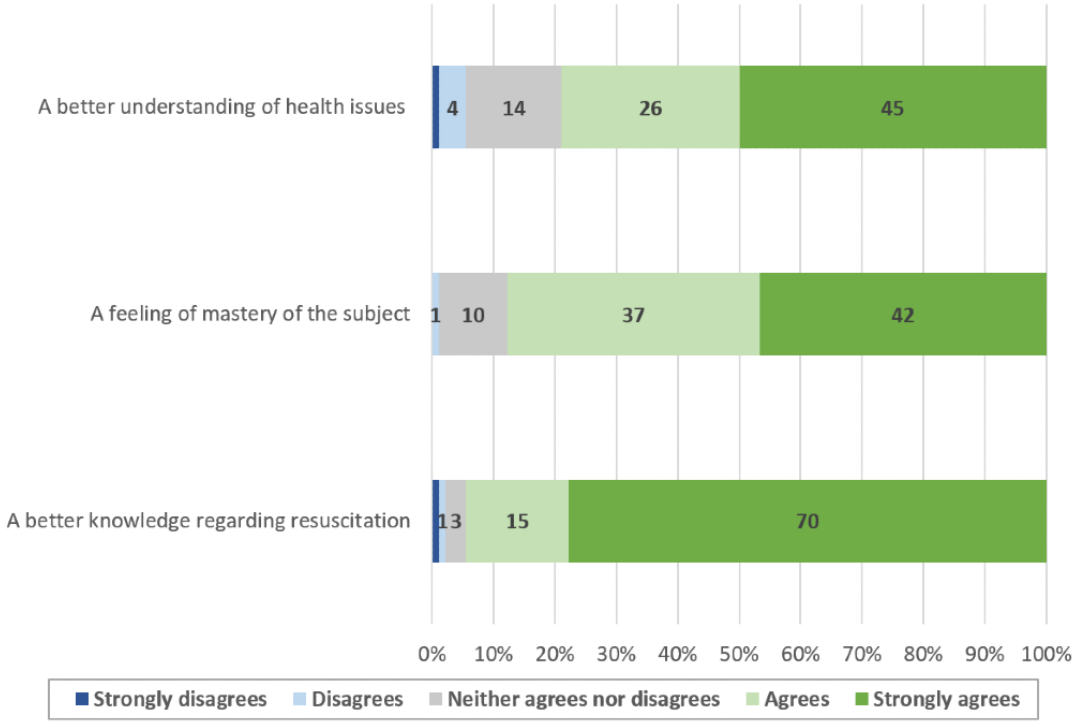
Characteristics	Enlisted as FR[a] (n=48)	Did not enlist as FR (n=114)	*P* value
Age (years), mean (SD)	20.3 (4.1)	20.1 (4.7)	.88
Gender, n (%)			.71
Male	14 (29.2)	30 (26.3)	
Female	34 (70.1)	84 (73.7)	
Other	0 (0)	0 (0)	
Curriculum, n (%)			.95
Medical	46 (95.8)	109 (95.6)	
Dental	2 (4.2)	5 (4.4)	
Knowledge score, mean (SD)			
10-point score	5.1 (1.7)	4.9 (1.6)	.56
6-point score	2.6 (1.0)	2.4 (1.1)	.19
Already followed a BLS[b] course, n (%)	38 (79.2)	80 (70.2)	.24
Wishes for more BLS training, n (%)	48 (100)	114 (100)	N/A[c]

[a]FR: first responder.

[b]BLS: basic life support.

[c]N/A: not applicable.

Figure 2. Factors contributing to an increased confidence in resuscitation skills.



XSL•FO
RenderX

Figure 3. Factors limiting student confidence in resuscitation skills. CPR: cardiopulmonary resuscitation.

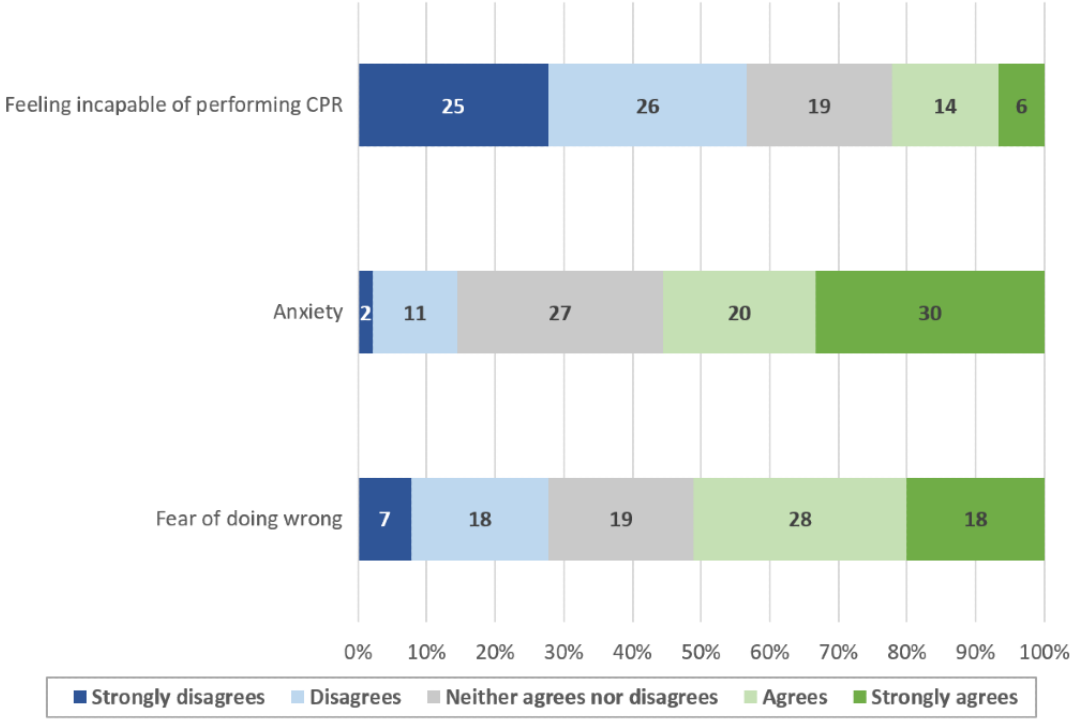
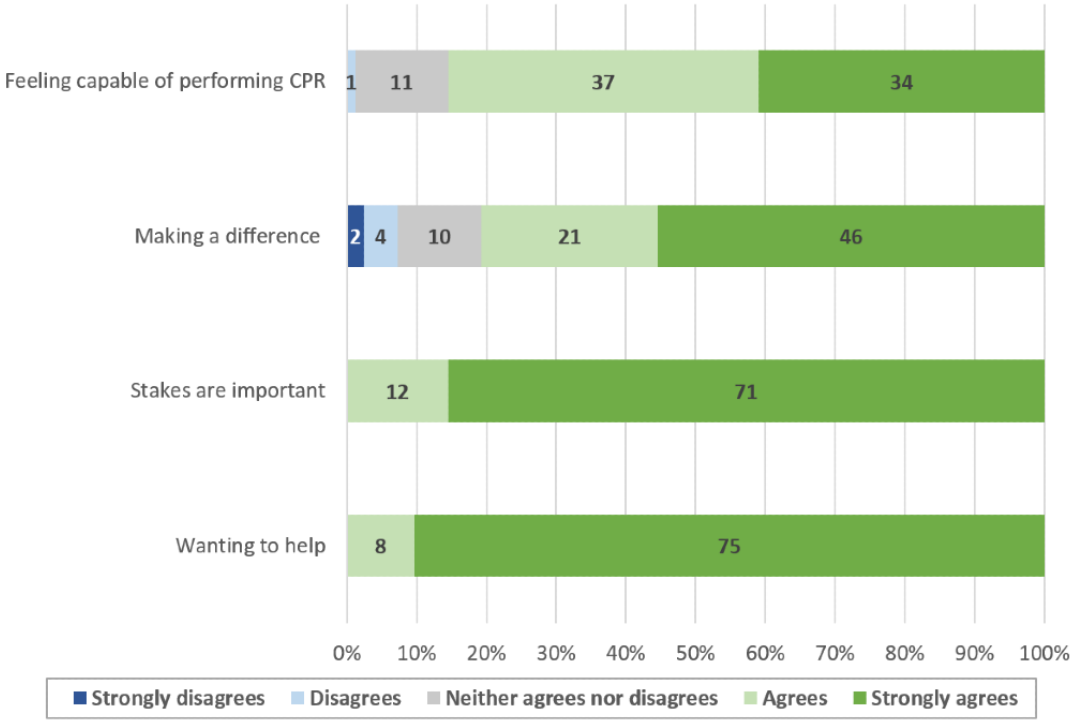
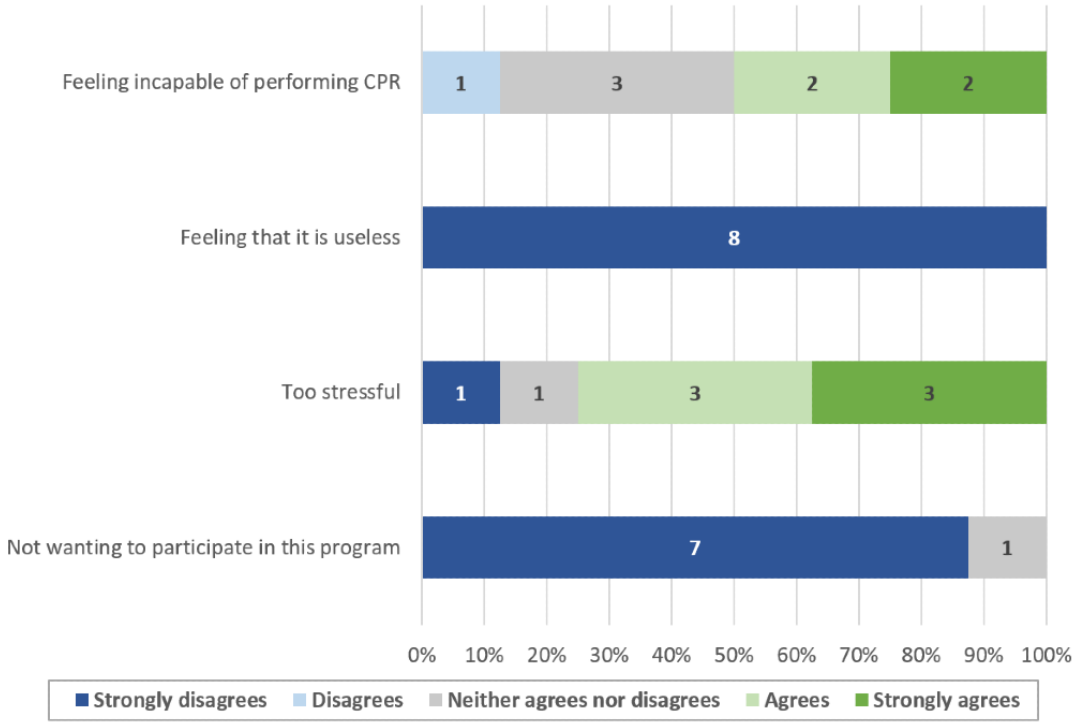


Figure 4. Factors motivating students to enlist as first responders. CPR: cardiopulmonary resuscitation.



XSL•FO
RenderX

Figure 5. Factors preventing students from enlisting as first responders. CPR: cardiopulmonary resuscitation.



Discussion

Main Considerations

This implementation study shows that less than 9.1% (n=48) of first-year UGFM medical and dental students enlisted as first responders after following a short motivational intervention and being offered the opportunity of following a blended learning path including an asynchronous e-learning module and hands-on practice. This proportion was slightly lower than the 10% target we had aimed at according to our original protocol [18].

Numerous reasons might explain this lower-than-expected result. First, even though the presentation was a success with almost half the promotion connecting to the platform, it took place rather late in the academic year, and practice sessions were scheduled even later. Therefore, because these junior students are under considerable pressure given the high failure rate, many might have elected to prioritize their revisions for the final exams over joining the first responder system regardless of their interest. Future studies will need to assess whether earlier inception can lead to higher registration rates. Second, while we strived to shorten the practice sessions in an attempt to increase the participation rate, some students might have felt that these sessions were too short to allow them to master BLS procedures. This might have prevented some participants from registering as first responders. Finally, a rather high proportion of students experienced technical difficulties with the registration component. Despite the help provided by email, this issue may have increased attrition at this stage.

The restrictions linked to the COVID-19 pandemic context may have played a role in our study. As all faculty courses were web-based, our first intervention was aired on the university streaming platform allowing students to skip it at will. This format also prevented us from interacting with the students and from answering their questions right away, and all further communications were carried out by email. While these elements might have contributed to a lower participation rate, the fact that we offered one of the few face-to-face courses might have played in our favor. Indeed, some comments showed that the students who participated suffered from a lack of social interaction and felt that the practice session was an almost unique occasion to discover the university premises and to directly talk to their peers. To assess the impact of these changes in the academic curriculum and given the relatively high enthusiasm showed by some students, this learning path will be offered once again during the academic year 2021-2022.

The confidence of the participants regarding their resuscitation skills was significantly improved, and the contributing factor most often reported by the students was a better knowledge of the subject. In our study, a higher confidence was associated with a higher probability of registering as first responder. However, less than two thirds of the students who completed the whole training process and obtained a course completion certificate had enlisted as first responders at 6 months. Interventions further strengthening student confidence in their resuscitation skills could help increase this proportion, and their impact should be assessed through further studies. Nevertheless, since people who feel more confident in their abilities are also more likely to act when faced with a stressful emergency [27], the participants who decided not to enlist as first responders might still be more likely to perform CPR if needed after following the learning path. This could prove particularly valuable since many studies show that BLS knowledge and skills are often limited in health care workers and students. Furthermore, starting resuscitation courses sooner during the medical curriculum could be advantageous since resuscitation skills tend to improve according to the number of BLS training

XSL•FO
RenderX

sessions [28,29]. In addition, the majority of students who reported wanting to enroll as first responders cited a desire to help as a main contributing factor. If the benefits of teaching BLS-AED courses during the first year of medicine curriculum could have an important impact on public health, generalizing this principle to mandatory school could have an even bigger one. Taking a step back from health care students, the European Resuscitation Council (ERC) recommends teaching CPR to children, preferably before the age of 12 years, in order to create long-lasting psychomotor skills and remember a short sequence of action over time. By doing so, targeting schoolchildren would be a way to increase the number of bystanders capable of performing CPR before the arrival of professional help and decrease the time of no flow [30].

In the past few years, the development of web-based courses has been expanding quickly, and the COVID-19 pandemic has increased this phenomenon even further [31]. Given the high potential number of interested students, the limited time available in the schedule of first-year medical and dental students, and the context of the COVID-19 pandemic, we believe that using an asynchronous distance e-learning module for the theoretical part of the course was particularly appropriate. Many studies have compared traditional didactic methods with e-learning interventions, the main limit being the huge variability of e-learning formats [32-34]. Among other parameters, interactivity and gamification mechanisms have been shown to affect the impact of e-learning interventions. Interactive e-learnings and serious games have been used in a few studies in Geneva and have demonstrated their superiority to the control group using usual course format [35,36]. While such modules have been shown to be useful to teach CPR procedures and decrease the time spent in workshops, they cannot replace hands-on practice [37].

Limitations

Apart from our main limitation (ie, the adaptation of our study to the pandemic context with the subsequent limited interaction with first-year UGFM students), other limitations must be acknowledged. First and foremost, our design did not include a comparator, the lack of which prevented us from determining whether a certain type of motivational intervention would have been more effective than another. In addition, the type of learning path could also influence the intention of registering as first responder. Nevertheless, this study was carried out in accordance with our original protocol, and the limited number of first-year UGFM students would have limited the power of our study had a comparator been included. Moreover, 6 months represent a rather long delay between the initial intervention and the assessment of the number of UGFM students who had enlisted as first responders. However, more than 3 months had elapsed between the initial intervention and the last practice sessions, and it is improbable that elements other than our learning path would have prompted medical students to enlist as first responders given the high workload associated with their end-of-year exams, which were scheduled less than 2 months after the last practice session had been completed.

There is also bias in the answers to the questionnaires as the students participating in the study were interested beforehand and could not fully represent the knowledge of the whole promotion. Furthermore, we asked the participants about their confidence during the postcourse questionnaire, which is a recollection bias, as the students may have had a misperception of their confidence, overestimating or underestimating their abilities.

Perspectives

Our initiative is in line with the concept of “systems saving lives” developed by the ERC [38]. Promoting and perpetuating such initiatives should help achieve the recommended number of available first responders, which is 10/km^2 according to ERC guidelines [38,39]. To further increase the awareness of medical and dental students regarding the importance of BLS-AED procedures, 2 studies regarding resuscitation skills of medical students will be conducted during the academic year 2021-2022 at UGFM. The first one will evaluate the impact of our intervention on BLS knowledge in second-year medical students. The second study will follow a very similar protocol as the present one and assess the effect of the intervention after addressing the weaknesses identified. Therefore, the questionnaires will be slightly modified, and the motivational intervention will take place sooner in the academic year to avoid interference with final revisions and exams. The effect of less-constraining COVID-19 restrictions on participation will be assessed in the course of this study.

Conclusion

After following a short motivational intervention, less than 10% of first-year medical and dental students enlisted as first responders after completing a blended learning path including an asynchronous e-learning module and hands-on practice. Including these future health care professionals in the first aid system early in their career and increasing the sheer number of potentially available first responders could help enhance survival and neurological outcomes in those having a cardiac arrest and participate in the building of their professional identity as a secondary benefit. Further studies are needed to understand the low proportion of medical and dental students finally enlisting as first responders and to determine whether different or additional teaching methods could increase this proportion.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Email sent to all first-year students.
[PDF File (Adobe PDF File), 1022 KB - jmir_v24i5e38508_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
Screening questionnaire and consent form.
[PDF File (Adobe PDF File), 486 KB - jmir_v24i5e38508_app2.pdf]

Multimedia Appendix 3
Disclaimer on the website's main page.
[PDF File (Adobe PDF File), 118 KB - jmir_v24i5e38508_app3.pdf]

Multimedia Appendix 4
Additional information regarding the study.
[PDF File (Adobe PDF File), 189 KB - jmir_v24i5e38508_app4.pdf]

Multimedia Appendix 5
Precourse questionnaire.
[PDF File (Adobe PDF File), 655 KB - jmir_v24i5e38508_app5.pdf]

Multimedia Appendix 6
Postcourse questionnaire.
[PDF File (Adobe PDF File), 509 KB - jmir_v24i5e38508_app6.pdf]

Multimedia Appendix 7
Original data (in CSV and DTA formats).
[ZIP File (Zip Archive), 52 KB - jmir_v24i5e38508_app7.zip]

References

1. 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 2022-05-11]
2. Gräsner J, Wnent J, Herlitz J, Perkins GD, Lefering R, Tjelmeland I, et al. Survival after out-of-hospital cardiac arrest in Europe - Results of the EuReCa TWO study. Resuscitation 2020 Mar 01;148:218-226. [doi: 10.1016/j.resuscitation.2019.12.042] [Medline: 32027980]
3. Yasunaga H, Horiguchi H, Tanabe S, Akahane M, Ogawa T, Koike S, et al. Collaborative effects of bystander-initiated cardiopulmonary resuscitation and prehospital advanced cardiac life support by physicians on survival of out-of-hospital cardiac arrest: a nationwide population-based observational study. Crit Care 2010;14(6):R199 [FREE Full text] [doi: 10.1186/cc9319] [Medline: 21050434]
4. Sturny L, Regard S, Larribau R, Niquille M, Savoldelli GL, Sarasin F, et al. Differences in Basic Life Support Knowledge Between Junior Medical Students and Lay People: Web-Based Questionnaire Study. J Med Internet Res 2021 Feb 23;23(2):e25125 [FREE Full text] [doi: 10.2196/25125] [Medline: 33620322]
5. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult basic life supportautomated external defibrillation section Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015 Oct;95:81-99. [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
6. Travers A, Perkins G, Berg R, Castren M, Considine J, Escalante R, et al. Part 3: Adult Basic Life Support and Automated External Defibrillation. Circulation 2015 Oct 20;132(16_suppl_1):A. [doi: 10.1161/cir.0000000000000272]
7. Larribau R, Deham H, Niquille M, Sarasin FP. Improvement of out-of-hospital cardiac arrest survival rate after implementation of the 2010 resuscitation guidelines. PLoS One 2018 Sep 24;13(9):e0204169 [FREE Full text] [doi: 10.1371/journal.pone.0204169] [Medline: 30248116]
8. Scquizzato T, Pallanch O, Belletti A, Frontera A, Cabrini L, Zangrillo A, et al. Enhancing citizens response to out-of-hospital cardiac arrest: A systematic review of mobile-phone systems to alert citizens as first responders. Resuscitation 2020 Jul;152:16-25 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.05.006] [Medline: 32437783]
9. Sasson C, Rogers MA, Dahl J, Kellermann AL. Predictors of Survival From Out-of-Hospital Cardiac Arrest. Circ: Cardiovascular Quality and Outcomes 2010 Jan;3(1):63-81. [doi: 10.1161/circoutcomes.109.889576]
10. Stiell IG, Wells GA, DeMaio VJ, Spaite DW, Field BJ, Munkley DP, et al. Modifiable factors associated with improved cardiac arrest survival in a multicenter basic life support/defibrillation system: OPALS Study Phase I results. Ontario Prehospital Advanced Life Support. Ann Emerg Med 1999 Jan;33(1):44-50. [doi: 10.1016/s0196-0644(99)70415-4] [Medline: 9867885]

XSL•FO
RenderX

11. Scquizzato T, Belloni O, Semeraro F, Greif R, Metelmann C, Landoni G, et al. Dispatching citizens as first responders to out-of-hospital cardiac arrests: a systematic review and meta-analysis. Eur J Emerg Med 2022 Jun 01;29(3):163-172. [doi: 10.1097/MEJ.0000000000000915] [Medline: 35283448]
12. Oving I, de Graaf C, Masterson S, Koster RW, Zwinderman AH, Stieglis R, ESCAPE-NET Investigators. European first responder systems and differences in return of spontaneous circulation and survival after out-of-hospital cardiac arrest: A study of registry cohorts. Lancet Reg Health Eur 2021 Feb;1:100004 [FREE Full text] [doi: 10.1016/j.lanepe.2020.100004] [Medline: 35104306]
13. Mauri R, Burkart R, Benvenuti C, Caputo ML, Moccetti T, Del Bufalo A, et al. Better management of out-of-hospital cardiac arrest increases survival rate and improves neurological outcome in the Swiss Canton Ticino. Europace 2016 Mar 07;18(3):398-404 [FREE Full text] [doi: 10.1093/europace/euv218] [Medline: 26346920]
14. Baldi E, Auricchio A, Klersy C, Burkart R, Benvenuti C, Vanetta C, SWISSRECA researchers. Out-of-hospital cardiac arrests and mortality in Swiss Cantons with high and low COVID-19 incidence: A nationwide analysis. Resusc Plus 2021 Jun;6:100105 [FREE Full text] [doi: 10.1016/j.resplu.2021.100105] [Medline: 34223367]
15. Sauvez une vie en vous engageant à nos côtés. Save a Life. URL: https://www.save-a-life.ch/ [accessed 2022-05-11]
16. Rapport d'activité 2019. Save a Life. 2020. URL: https://www.save-a-life.ch/wp-content/uploads/2021/05/save-a-life-rapport-dactivit-2019.pdf [accessed 2022-05-11]
17. Xie JY, Frost R, Meakin R. Not quite a doctor, but should I help? A qualitative exploration of medical students' attitudes towards responding to medical emergencies that occur in the public domain. BMJ Open 2019 Apr 08;9(4):e028035 [FREE Full text] [doi: 10.1136/bmjopen-2018-028035] [Medline: 30962243]
18. Suppan L, Herren T, Taramarcaz V, Regard S, Martin-Achard S, Zamberg I, et al. A Short Intervention Followed by an Interactive E-Learning Module to Motivate Medical Students to Enlist as First Responders: Protocol for a Prospective Implementation Study. JMIR Res Protoc 2020 Nov 06;9(11):e24664 [FREE Full text] [doi: 10.2196/24664] [Medline: 33155574]
19. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Federal Act on Research involving Human Beings (Human Research Act, HRA). The Federal Assembly of the Swiss Confederation. 2011 Nov 30. URL: https://www.admin.ch/opc/en/classified-compilation/20061313/index.html [accessed 2022-05-11]
21. Joomla Content Management System. Joomla. URL: https://www.joomla.org/about-joomla.html [accessed 2022-05-11]
22. General Data Protection Regulation (GDPR) Compliance Guidelines. General Data Protection Regulation. URL: https://gdpr.eu/ [accessed 2020-09-30]
23. Directives de cours 2015 Guide pour l'obtention du label SRC pour les cours Basic Life Support. Swiss Resuscitation Council. 2015. URL: https://www.resuscitation.ch/fileadmin/user_upload/Kursrichtlinien_2015_franz.pdf [accessed 2022-05-11]
24. Nolan JP, Monsieurs KG, Bossaert L, Böttiger BW, Greif R, Lott C, European Resuscitation Council COVID-Guideline Writing Groups. European Resuscitation Council COVID-19 guidelines executive summary. Resuscitation 2020 Aug;153:45-55 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.06.001] [Medline: 32525022]
25. Modification des mesures nationales visant à endiguer le coronavirus en Suisse depuis décembre 2020. Federal Office of Public Health of the Swiss Confederation. 2021. URL: https://www.bag.admin.ch/dam/bag/fr/dokumente/mt/k-und-i/aktuelle-ausbrueche-pandemien/2019-nCoV/tabelle-aenderungen-massnahmen.pdf.download.pdf/Modification_mesures.pdf [accessed 2022-05-11]
26. Federal Office of Public Health of the Swiss Confederation. Assouplissement et renforcement des mesures nationales. Bulletin des médecins suisses 2015 Jan 27;96(05):1-5. [doi: 10.4414/bms.2015.03321]
27. Shida H, Nishiyama C, Okabayashi S, Yamamoto Y, Shimamoto T, Kawamura T, et al. Laypersons' Psychological Barriers Against Rescue Actions in Emergency Situations - A Questionnaire Survey. Circ J 2022 Mar 25;86(4):679-686 [FREE Full text] [doi: 10.1253/circj.CJ-21-0341] [Medline: 34759132]
28. Grześkowiak M. The effects of teaching basic cardiopulmonary resuscitation--a comparison between first and sixth year medical students. Resuscitation 2006 Mar;68(3):391-397. [doi: 10.1016/j.resuscitation.2005.07.017] [Medline: 16457932]
29. Niemi-Murola L, Mäkinen M, Castren M, ECCE Study Group. Medical and nursing students' attitudes toward cardiopulmonary resuscitation and current practice guidelines. Resuscitation 2007 Feb;72(2):257-263. [doi: 10.1016/j.resuscitation.2006.07.006] [Medline: 17126983]
30. Böttiger BW, Lockey A, Georgiou M, Greif R, Monsieurs KG, Mpotos N, et al. KIDS SAVE LIVES: ERC Position statement on schoolteachers' education and qualification in resuscitation. Resuscitation 2020 Jun;151:87-90 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.04.021] [Medline: 32339597]
31. Rose S. Medical Student Education in the Time of COVID-19. JAMA 2020 Jun 02;323(21):2131-2132. [doi: 10.1001/jama.2020.5227] [Medline: 32232420]
32. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]

XSL•FO
RenderX

33. van Gaalen AEJ, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma ADC, Georgiadis JR. Gamification of health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2021 May;26(2):683-711 [FREE Full text] [doi: 10.1007/s10459-020-10000-3] [Medline: 33128662]
34. Wheeler S. e-Learning and Digital Learning. Encycl Sci Learn 2012:1111. [doi: 10.1007/978-1-4419-1428-6_431]
35. Koka A, Suppan L, Cottet P, Carrera E, Stuby L, Suppan M. Teaching the National Institutes of Health Stroke Scale to Paramedics (E-Learning vs Video): Randomized Controlled Trial. J Med Internet Res 2020 Jun 09;22(6):e18358 [FREE Full text] [doi: 10.2196/18358] [Medline: 32299792]
36. Suppan M, Gartner B, Golay E, Stuby L, White M, Cottet P, et al. Teaching Adequate Prehospital Use of Personal Protective Equipment During the COVID-19 Pandemic: Development of a Gamified e-Learning Module. JMIR Serious Games 2020 Jun 12;8(2):e20173 [FREE Full text] [doi: 10.2196/20173] [Medline: 32516115]
37. Boada I, Rodriguez-Benitez A, Garcia-Gonzalez JM, Olivet J, Carreras V, Sbert M. Using a serious game to complement CPR instruction in a nurse faculty. Comput Methods Programs Biomed 2015 Nov;122(2):282-291. [doi: 10.1016/j.cmpb.2015.08.006] [Medline: 26319184]
38. Semeraro F, Greif R, Böttiger BW, Burkart R, Cimpoesu D, Georgiou M, et al. European Resuscitation Council Guidelines 2021: Systems saving lives. Resuscitation 2021 Apr;161:80-97. [doi: 10.1016/j.resuscitation.2021.02.008] [Medline: 33773834]
39. Stieglis R, Zijlstra JA, Riedijk F, Smeekes M, van der Worp WE, Koster RW. AED and text message responders density in residential areas for rapid response in out-of-hospital cardiac arrest. Resuscitation 2020 May;150:170-177 [FREE Full text] [doi: 10.1016/j.resuscitation.2020.01.031] [Medline: 32045663]

Abbreviations

AED: automatic external defibrillator
BLS: basic life support
CPR: cardiopulmonary resuscitation
ERC: European Resuscitation Council
OHCA: out-of-hospital cardiac arrest
UGFM: University of Geneva, Faculty of Medicine

Edited by A Mavragani; submitted 06.04.22; peer-reviewed by T Shimamoto, S El kefi, T Scquizzato; comments to author 25.04.22; revised version received 25.04.22; accepted 28.04.22; published 18.05.22.

Please cite as:
Taramarcaz V, Herren T, Golay E, Regard S, Martin-Achard S, Mach F, Schnetzler N, Ricci G, Zamberg I, Larribau R, Niquille M, Suppan M, Schiffer E, Suppan L
A Short Intervention and an Interactive e-Learning Module to Motivate Medical and Dental Students to Enlist as First Responders: Implementation Study
J Med Internet Res 2022;24(5):e38508
URL: https://www.jmir.org/2022/5/e38508
doi:10.2196/38508
PMID:35583927

©Victor Taramarcaz, Tara Herren, Eric Golay, Simon Regard, Sébastien Martin-Achard, Francois Mach, Nicolas Schnetzler, Gaëtan Ricci, Ido Zamberg, Robert Larribau, Marc Niquille, Mélanie Suppan, Eduardo Schiffer, Laurent Suppan. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.05.2022. 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.

XSL•FO
RenderX

Original Paper

Exploring Online Peer Support Groups for Adults Experiencing Long COVID in the United Kingdom: Qualitative Interview Study

Hannah L S Day[1], MSc

Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Hannah L S Day, MSc
Faculty of Public Health and Policy
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 2079272700
Email: hannah.day1@alumni.lshtm.ac.uk

Abstract

Background: Long COVID is an emerging public health concern. A growing number of individuals are experiencing prolonged, multifaceted health challenges and accompanying social impacts after COVID-19 infections. Support services in the United Kingdom remain insufficient and fraught with complexity. Responding to persistent gaps in care, patients joined forces in online peer support groups. However, little is known about how these groups impact patients with long COVID and their lived experiences of the condition.

Objective: The aim of this study is to explore the roles that online peer support groups take on and the impact they have on patients experiencing and recovering from long COVID in the United Kingdom. In doing so, this study aims to identify ways to inform future long COVID care, including online peer support and broader long COVID care structures.

Methods: I conducted 11 semistructured interviews virtually on Zoom in July 2021. Participants had long COVID, were UK-based, and used long COVID online peer support groups. Topics discussed in interviews included what led participants to these groups, experiences within them, and feelings about the roles that the groups took on. I analyzed the results by manually conducting thematic analysis.

Results: Long COVID online peer support groups had numerous roles, significantly impacting users. I identified 5 themes and 13 subthemes through thematic analysis. The identified themes were as follows: (1) filling professional care gaps, (2) societal awareness, (3) engagement behavior, (4) diversity, and (5) social connections. Given the void of professional support, those experiencing long COVID gained some benefit from these groups. However, participants emphasized notable concerns about the all-encompassing roles these groups embody and speculated over potential improvements.

Conclusions: If used appropriately, online peer support groups could be immensely beneficial for patient well-being, beyond simply filling gaps in long COVID care. However, it appears many groups take on more than they can manage and become potentially harmful. Through prioritizing patient voices, long COVID care could be restructured to maximize peer support's benefits within broader care structures.

(J Med Internet Res 2022;24(5):e37674) doi:10.2196/37674

KEYWORDS

COVID-19; long COVID; post-COVID-19 syndrome; peer support; online health communities; self-help groups; internet; qualitative; interview; patient experience; digital health; digital peer support; online health

XSL•FO
RenderX

Introduction

Context of Long COVID

Prevalence and Impact

The COVID-19 pandemic has significantly impacted the UK population. As of March 29, 2022, there had been 20,986,171 confirmed COVID-19 cases and 164,974 deaths within 28 days of a positive COVID-19 test [1,2]. However, as Tim Spector (creator of the ZOE COVID Symptom Study app) noted in May 2020, there is a misconception that "if you are not dead you are fine," thus leaving continued suffering unacknowledged [3].

Long COVID encapsulates both "ongoing symptomatic COVID-19" (symptoms experienced for 4-12 weeks from infection) and "post-COVID-19 syndrome" (symptoms 12 or more weeks from infection) [4]. Prevalence estimates in the United Kingdom from a survey completed on March 5, 2022, indicated that 1.7 million people in private households were actively experiencing long COVID symptoms without an alternative explanation [5]. An estimated 784,000 (45%) of those reporting long COVID stated they had COVID-19 over a year prior, and 74,000 (4%) reported it had been at least 2 years since their COVID-19 infection [5]. However, long COVID prevalence estimates vary widely and largely depend on self-reported data [6,7]. A study from February 2021 found that roughly 30% of patients with COVID-19 experienced long-lasting symptoms, even 9 months after infection [8]. This uncertainty is amplified by the widespread unavailability of COVID-19 testing early on. Long COVID may not be considered without confirmed COVID-19 infection [9]. With England's "Living with COVID-19" plan removing access to universal free COVID-19 tests from April 1, 2022 [10], long COVID's impacts may be further clouded as many will not have a positive test to document a prior infection.

Long COVID symptoms appear systemic, impacting multiple bodily systems with varying severity [4,11,12]. A survey uncovered 203 long COVID symptoms [11]. Common symptoms include immense fatigue, cognitive dysfunction ("brain fog"), palpitations, peripheral neuropathy, depression, breathing difficulties, autonomic dysfunction, and new-onset allergies [6,11,13]. This incomplete list mirrors long COVID's all-encompassing nature, unsettling assumptions that long COVID can be simplified into a single unidirectional illness trajectory.

In an Office for National Statistics survey in June 2021, roughly 57% of those self-reporting long COVID reported negative impacts on their well-being and 30% reported negative impacts on work [14]. Compared to those not experiencing long COVID (or any COVID-19 infection), those with long COVID fared worse across numerous indicators, such as anxiety and loneliness [14]. Long COVID does not impact everyone equally [5,6,13]. Women appear twice as likely to experience long COVID than men [15]. The 35-49-year age group was most likely to report long COVID [5], with increased risk for lower-income groups [6]. Preexisting disability and health conditions appear to increase risk [5], particularly existing lung conditions [6]. However, current data should be approached with caution as research is evolving.

Definitions and Medicalization

The term "long COVID" was created as a hashtag (#LongCovid) by Elisa Perego in May 2020, naming her turbulent COVID-19 experience [16,17]. This phrase's popularity grew quickly, shifting into news and research [16]. Notably, there remains uncertainty around long COVID's definitions. The World Health Organization defines long COVID as symptoms lasting more than 3 months from initial COVID-19 infection [18]. However, the definition from the National Institute for Health and Care Excellence (NICE) used before expands this timeline to symptoms lasting more than 4 weeks [4]. Given that this research focuses on the UK population and corresponding data, I am utilizing NICE's definition. Additionally, this broader definition avoids unduly excluding patients who require essential services, particularly while long COVID remains poorly understood. Patients had to fight for long COVID recognition, medicalizing the condition. Historically, experts drove medicalization: situating natural experiences within biomedicine's purview [19,20]. Recently, patients have driven this phenomenon [21], as with long COVID. Long COVID is speculated as being the first illness defined solely through patients' social media communications [9,16]. Patients achieved validity through "illness reification," defining suffering through shared experience [21], thus accessing medical legitimization and care.

Existing Health Care Response

The National Health Service (NHS) has attempted to provide long COVID support. In the summer of 2020, NHS England launched "Your COVID Recovery," a website connecting patients with health care providers (HCPs) primarily advising self-management [22]. In October 2020, the NHS announced "post-COVID assessment clinics" [22]. The latest data indicate there are 90 specialist clinics in England [23]. There appear to be no clinics currently in Scotland or Wales [24-26]. Northern Ireland only announced specialist clinics in November 2021 [27]. Patient experience reflects these gaps, exacerbated by HCP dismissal and geographical variations in care ("the postcode lottery") [9,12,28-30]. In June 2021, the NHS published "Long COVID: The NHS Plan for 2021/22" [22]. However, the real-world rollout and impacts of this plan are not yet known.

Role of Peer Support

Given care gaps, many long COVID online peer support groups emerged on social media [31]. Peer-led interventions are multifaceted and often used for the management and prevention of various conditions, including HIV/AIDS, diabetes, and adjustment to chronic illness [32,33]. Increasingly, individuals seek medical advice and support online [21,34,35]. In 2012, Ziebland and Wyke identified 7 ways online sources impact patient experiences positively and negatively, including around information, support, health care usage, and creating illness narratives [35]. Although the online landscape has changed substantially due to social media's rapid recent growth since Ziebland and Wyke conducted their research, it provides strong foundations for this study.

XSL•FO
RenderX

Importance of This Study

Long COVID research remains sparse [12], focusing on symptomology and social impacts. However, research into support strategies' impacts is lacking. This study aims to address these epistemological gaps. There is temporal importance: as more people get COVID-19, more will experience long COVID. Cumulative burdens on health care accentuates the need to explore support possibilities.

This study's aim is to explore the role of online peer support groups in UK adults' recovery from long COVID, focusing on 2 objectives. The primary objective is to explore the impact these groups have on patients and their experiences within these spaces. The secondary objective is to identify ways these online peer support groups can be situated within broader long COVID recovery planning.

Methods

Study Design

I chose qualitative methods to generate richer data, unpacking participants' experience [36]. Semistructured interviews afforded flexibility; participant responses could shape the interview trajectory [36], while permitting a topic guide (Multimedia Appendix 1). I utilized a phenomenological approach in order to prioritize elucidation of participants' lived experiences and unpacking of complexities [37,38].

I aimed to recruit 10-12 participants. The decision surrounding sample size was the result of practical constraints, including a condensed time frame for my master's thesis schedule and burdens placed on me as a researcher with long COVID conducting this research individually. I contacted administrators from 1 Slack and 7 Facebook groups containing over 7500 members for permission to post my recruitment poster, ensuring the post's appropriateness [39]. Of those, 3 (43%) Facebook groups and the Slack group allowed me to post my recruitment poster. I also shared the poster on my social media to increase exposure.

I used convenience sampling, selecting participants on a first-come, first-served basis. I successfully recruited 11 participants before closing recruitment on July 19, 2021. I only included UK-based adults. Adults have greater control over health-related decisions, and it allowed for focused discussion within the UK health care context.

Data Collection

I conducted 11 interviews between July 14 and 27, 2021, lasting 26-78 minutes (averaging 49.5 minutes). All participants were offered 2 interviews to ensure they felt their narrative was fully heard, although all opted for 1. I conducted interviews using Zoom, a practical alternative for remote data collection [40,41]. Although initially chosen due to the pandemic, Zoom interviews provided notable advantages. Severe long COVID may preclude travel, reducing participant diversity [41]. It also allowed me to reach people beyond London where support may be less concentrated. Additionally, videoconferencing facilitated more rapport compared to telephone interviews [41]. Only 1 participant kept their video off; however, I kept mine on so they could see my reactions.

I recorded the interviews through Zoom, manually transcribing them on Microsoft Word aided by Express Scribe and an external foot pedal. I transcribed verbatim, reproducing how words were spoken using Poland's abbreviations [42]. This notation style includes the following: short pauses indicated with dots in parentheses, for example, "(.)" or "(..)"; relaying speech from others and sharing the internal narrative indicated by "(mimicking voice)"; and overlapping speech indicated by a hyphen when the interjection occurs, and the dialogue of the second speaker begins with "(overlapping)" [42]. I cross-checked transcripts against original recordings to ensure validity. I chose to not amend the grammar in participants' speech in either the transcripts or the writeup. I aimed to capture how words were spoken without "cleaning" the speech with my own biased, uniquely molded, conversational, and linguistic paradigm. To aid clarity in writeup, I used "[…]" to indicate omitted dialogue.

Data Analysis

After data collection, I utilized reflexive thematic analysis to deeply explore experiences within long COVID online peer support groups that have not yet been formally brought to light. I approached analysis with a constructivist epistemology, stipulating that knowledge is generated through social constructions and variable interpretations [43].

Using Braun and Clarke's guidance [44], I conducted thematic analysis manually. I coded each transcript inductively—data driven, not using predetermined code lists [36,44]. From my initial code list, I formed related categories using Microsoft Excel. These categories were then grouped into potential themes and subthemes. I refined potential themes, identifying those related specifically to this research [44]. Using Excel, I organized all coded data extracts into each subtheme. I saved documents to provide a comprehensive audit trail of my decision-making throughout the coding process.

Additionally, as I was the sole researcher and interpreter of the data, I engaged in peer debriefing to improve this study's credibility [45,46]. Peer debriefing is the "process of exposing oneself to a disinterested peer in a manner paralleling an analytic session and for the purpose of exploring aspects of inquiry that might otherwise remain only implicit within the inquirer's mind" [45]. I chose a peer for this process who is a former colleague with experience in qualitative public health research and with whom I have a relationship built on honesty and trust [47]. We engaged in this process in the final stages of data analysis and reporting to provide additional perspectives on my codes, increase my awareness of any oversights or biases entering the analysis, and troubleshoot redefining my themes.

Ethical Considerations

The London School of Hygiene & Tropical Medicine's MSc Research Ethics Committee granted ethics approval for this study (reference numbers 25478, 25478-1, and 25478-2).

The study conformed to the ethical principles of the Declaration of Helsinki. Participants received an information sheet

XSL•FO
RenderX

(Multimedia Appendix 2) containing comprehensive study information and my contact details. Following sufficient time for questions or clarification, I obtained informed consent from all participants. Forms, recordings, and transcripts were stored securely. Additionally, I used pseudonyms to protect each individual's identity owing to the small number of participants in the study.

All efforts were made to ensure interviews were private, though this could not be guaranteed with participants joining Zoom calls from their chosen locations (often, their homes). I was in a private room with headphones and began interviews by enquiring about the risk of their privacy being compromised. If there was a risk, I planned to collaboratively create a code word [48]. Use of this word would have facilitated a conversational shift and prompt friendly close.

Moreover, as participants in my study all experienced long COVID, some may have found interviews exhausting. I reassured them they could manage symptoms, as needed, including ending interviews early. Additionally, interviews could have emotional impacts [36,49], particularly if participants struggled to obtain support in their long COVID journey. I sent participants a document for additional support and guidance, using resources available at the time (Multimedia Appendix 3). Lastly, I checked in with participants to ensure they were not feeling distressed as we concluded the interviews [49].

Reflexivity and Positionality

As a young adult woman with preexisting health conditions and long COVID following COVID-19 in March 2020, I embodied both researcher and patient. My positionality uniquely impacted the coproduction of knowledge [50,51], providing valuable depth. As an insider, I had the benefit of acute awareness of the issue and understanding language used by participants in describing their experience [52]. Of course, there was a risk of my own biases in analysis and interpretation, such as overseeing data that I take for granted [53,54] or inadvertently allowing my experiences to influence my approach [54]. I avoided these issues by ensuring I considered possible biases in advance and was self-reflexive throughout. Additionally, in engaging in peer debriefing with someone who did not have long COVID, the risk of any subconscious biases influencing data analysis was reduced [45].

I embodied my illness by sitting on the floor during interviews and disclosing it at multiple times. This openness strengthened my connection with participants and broke down research's traditional unequal power dynamics [51,55]. In being open, participants could ask about my long COVID journey [55]. These questions allowed for a more natural conversational environment, though they shifted conversation off track at times. Additionally, I found myself internalizing the intense emotions [55], particularly when a participant's journey mirrored my own. I had to create a space where I could step back after interviews to protect my well-being.

Results

Summary

From the interviews, I identified 5 themes and 13 subthemes (see Figure 1). Participants, whose characteristics are described in Table 1, had varying long COVID experiences. In total, 10 (91%) of 11 participants used Facebook for their online support (each with over 7500 members); Jessica used a small WhatsApp group (of roughly 22 group members).

The analysis revealed overarching commonalities between participants' reflections of these long COVID online peer support groups, which was well articulated by James: "[…] it's something that's there, but it's not what's needed." The intricacies of their experience within these groups and the roles groups take on are unpacked in the upcoming themes.

XSL•FO
RenderX

Figure 1. Conceptual thematic map. The hierarchy of concepts identified in thematic analysis are represented here: overarching concepts (bolded and underlined text boxes), themes (bolded text boxes), and subthemes (plain text boxes).

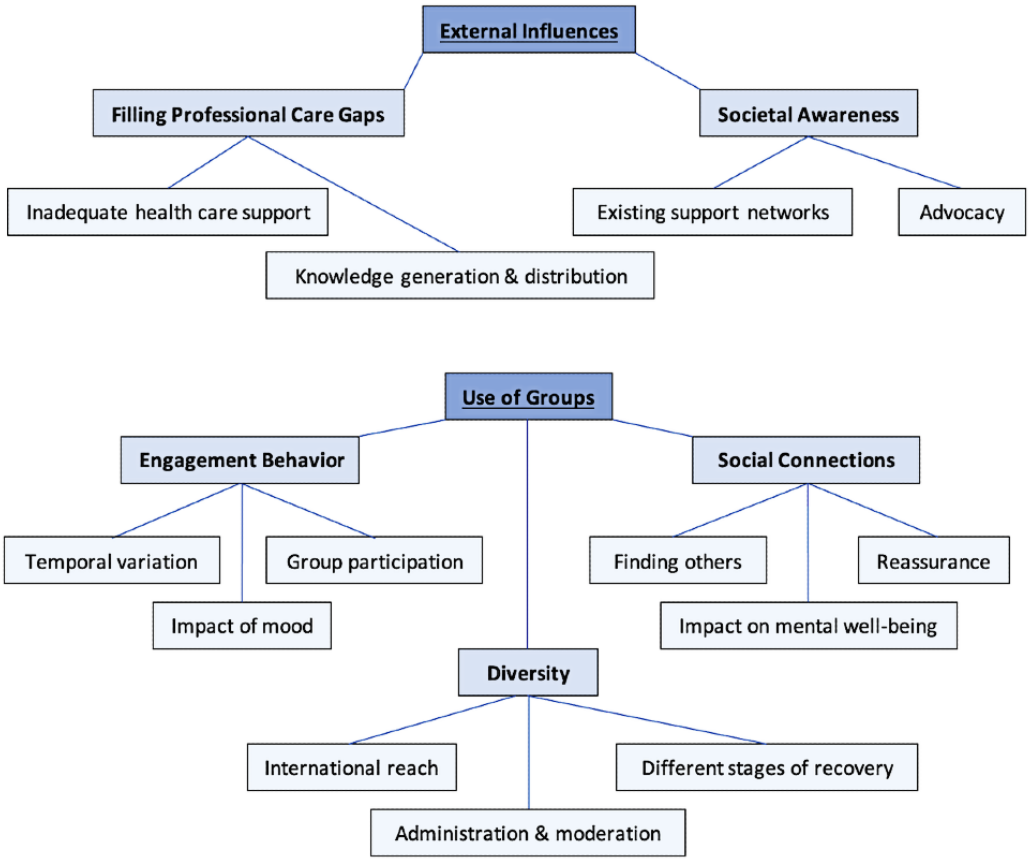


Table 1. Participant characteristics (N=11).

Participant (pseudonym)	Age group (years)	Participant-identified gender	Date of (suspected) COVID-19	When they joined the support group(s) (approx)	Support group membership (approx), n
Chloe	30-39	Woman	September 2020	October 2020	2
James	30-39	Man	June 2020	January 2021	1
Jessica	40-49	Woman	May 2020	August 2020	1
Claire	50-59	Woman	February 2021	March 2021	3
Emma	30-39	Woman	October 2020	January-March 2021	3
Oliver	30-39	Man	March 2020	February-March 2021	3-4
Mia	30-39	Woman	January 2021	April 2021	3
Emily	20-29	Woman	January 2021	April 2021	3
Natalie	30-39	Woman	March 2020	June 2021	1
Will	50-59	Man	February 2021	March 2021	1
Sophia	30-39	Woman	October 2020	March 2021	1

Theme 1: Filling Professional Care Gaps

Inadequate Health Care Support

Central to many participants' narratives was how they turned to online peer support groups after facing an "abyss of silence" [Natalie] from HCPs, feeling let down by health systems.

> *[…] she said (mimicking voice) I think you probably do have long COVID. And that was like full stop. I kind of paused waiting for some - there was nothing.* [Natalie]

Although there are (limited) NHS long COVID clinics, participants reported challenges accessing them. For some, there were no clinics nearby that HCPs could refer them to; for others, access remained fraught with complexity.

> *I've been refused access for being too unwell (.) for the COVID clinic, that they say I need a (.) respiratory referral. But I was refused a respiratory referral*

XSL•FO
RenderX

> *because I was too well. […] I'm just sat in the middle.* [James]

The oversight on long COVID care led participants to compare to services offered for other conditions. James noted how with HIV, "at the end of the day there is always somewhere for someone to turn." Natalie further critiqued the lapses in care more broadly, noting that "often charities and support groups end up filling all the gaps."

HCPs (gatekeepers to further care) frequently dismissed concerns as mental health issues; patients were "getting gaslighted for that" [Oliver]. Others reported how HCPs repeatedly overlooked otherwise concerning symptoms, opting not to investigate the cause and instead stating, "(mimicking voice) it's long COVID you just need to (.) um recover" [Chloe]. This lack of adequate attention frustrated participants who were forced to take control over their recovery.

On the groups, some participants wanted more HCP involvement for moderation and "to like answer questions" [Emily]. However, others were content with limited HCP interference, being worried that medical presence "would take away from what the group actually is, which is peer support, isn't it?" [Mia]. Jessica's experience on the small WhatsApp group reflected a similar sentiment to Mia's:

> *Researcher: And are these doctors and medical professionals - are they part of the group at all? Or is the group just the patients recovering?*
>
> *Jessica: Um (.) so originally they're the ones that started it off originally. And then once everybody was sort of introduced, they left and you know - it was so that we could talk about things that were worrying us, without somebody from the medical field […].*

Knowledge Generation and Distribution

Owing to insufficient medical care, online peer support groups became spaces to share therapies and management techniques. Both Oliver and Natalie used similar language, highlighting the gaps groups aimed to fill:

> *At the moment we're all kind of swinging in the dark and kind of hoping to find something […].* [Oliver]

Mia noted how what was shared in these groups was "actually ahead of the medical information," coming from patient expertise. Many participants reported gratitude there was advice available, primarily beneficial when no alternatives existed.

However, some were concerned over the lack of content control, meaning they "just take everything with a pinch of salt" [Oliver]. There was palpable frustration regarding potential implications of unregulated content, particularly on social media where "it can be hard to-to differentiate between what is a sensible piece or post and what is a post that's maybe got ulterior motives to it" [Will].

> *[…] people might be trying to be helpful, but I think it's dangerous to try and be too helpful. Because you […] don't know what their symptoms are, you don't know what their situation is, you shouldn't be saying take this take that, because (.) someone could do something stupid and kill themselves.* [Will]

To avoid risks, several participants used information to signpost HCP discussions, with HCPs as "their safety net" [Sophia]. Although Mia noted risks of overprioritizing medical knowledge:

> *But then again, (sigh) because there's a lack of evidence base, there's an argument there for who is the health professional then?* [Mia]

This contentious relationship between experts and participants in these groups was tangible.

Theme 2: Societal Awareness

Existing Support Networks

It was clear that discussing long COVID with others was challenging. A few participants expressed fears of burdening loved ones.

> *[…] my daughter would go (mimicking voice) you alright? And I'd go nope! And I'd just start crying. And that's when I thought (.) I can't keep putting everything on her […].* [Claire]

Others reported feeling friends were not understanding, tiring of the topic, or underplaying the issues. For Sophia, leaving the house to vote was viewed differently by those without long COVID:

> *That felt like maybe the biggest thing I'd done in a week or a month or something. But (.) to my friend group that aren't (.) like (.) COVID sufferers or aren't really aware of how bad it is, they sort of read that and go (mimicking voice) really? That's an achievement?*

Numerous participants reported that support from others with long COVID provided different, more attuned, support:

> *[…] when it comes from people who have gone through the same thing, it feels a little bit more as in it's realistic and it will happen.* [Jessica]

As a result, these peer support groups provided notable advantages to participants' existing support networks.

Importantly, online support groups also provided spaces for loved ones to better understand long COVID from patients' perspectives. James' wife used the space "because I think she was getting a bit frustrated in that (.) (deep inhale) that there is nothing out there that you can just kind of, pick up and read about" [James].

Advocacy

Simply the existence of groups validated the condition's importance. Chloe stated:

> *I think the groups bringing it to light, that there is so many thousand people feeling this way means that someone's gonna have to step up and do something.*

Additionally, groups facilitated petition sharing and encouraged research, which was deeply valued by participants.

XSL•FO
RenderX

> *[…] thank God we've got people, you know, fighting for us and pushing this research forwards. Because it is an important topic […].* [Mia]

However, participants felt that groups had limited scope for impact due to societal power dynamics:

> *[…] on the Facebook groups it is literally Joe Bloggs, it's no (.) people that have power to change things?* [Oliver]

When considering where power truly lies, some participants wished for greater government action to provide adequate services, including to have a hand in support groups.

> *I almost think the Department of Health needs to take a little bit of (..) um control, […] things like the COVID clinics was just - they were completely underfunded before they even start. […] Um (…) we need some sort of, I don't know, register is probably not the right thing, but some way of identifying people - um (.) traffic-lighting them into support groups. The right support groups for the right people.* [James]

Theme 3: Engagement Behavior

Temporal Variation

Engagement in the online peer support groups was fluid. Participants joined groups at different times, elucidating different expectations of perceived benefit.

> *[…] it's sometimes too soon to be reaching out because I know I didn't do anything on the group until probably Christmas time. Because I just wanted to see how I would go. […] Whereas I think if you're too invested (..) you're never gonna feel like yourself again if you're constantly reading everyone else feeling miserable.* [Chloe]

For some, engagement changed over time: "[…] it's tapered off" [Jessica], they "dip into every now and then" [Will], or, like Oliver, they engaged differently after accessing long COVID clinics.

Group Participation

Participants engaged depending on their individual journeys, particularly if they felt their "experiences might be relevant to the question they've asked" [Will] or if they "can give some value" to questions asked [Chloe]. Indeed, there was a clear desire to give back to others, either through applying their professional expertise or in offering support.

> *[…] if I read it and I've taken the time to read it, and it's something that's resonated with me, I'll always try and comment back to them and just give them back a little bit of the support that I've found from posting on there.* [Sophia]

Impact of Mood

Notably, engagement behavior depended on mood and varied immensely between participants. Some participants used these groups when feeling low:

> *Because (.) at that point I think you need a wee boost.* [Chloe]

However, others actively avoided seeing group content in these moments:

> *[…] on a day when I'm not feeling great I probably avoid looking at it? Um, as much as possible.* [Natalie]

This discrepancy created an interesting tension, further highlighting how participants held different expectations of the groups' utility.

Theme 4: Diversity

International Reach

Participants in Facebook groups all noted the international catchment. Several participants highlighted this diversity's benefits:

> *I thought it'd be good to hear from other medical systems as well, maybe they've got other ideas or different ways of tackling this.* [Natalie]

However, others expressed potential pitfalls of this global reach, including different terminology causing confusion, overwhelming information, and confronting more stories of suffering. Mia expressed this commonly held sentiment:

> *[…] sometimes when you're enlightened to other people's struggles in other countries, when you're unwell yourself mentally, it can have quite a big impact I think.*

Different Stages of Recovery

Groups contained people at all stages of long COVID, with widely varying symptoms and experiences. Many participants compared their experience to others, eliciting a spectrum of emotions. When comparing durations of their illness, there was palpable anxiety and even hopelessness for potential prolonged suffering from reading others' tumultuous journeys.

> *[…] people aren't being negative, and they're just saying how long their journey is, that to me I think, well I don't think I can do this for another like (.) 6 months.* [Emma]

> *Sometimes it worries me a bit because there's people on there who've had these symptoms for 18 months, and I'm thinking oh my God, please no.* [Mia]

Others expressed frustration when group members complained about comparatively shorter long COVID journeys.

> *[…] when you get people that have, that have lost their taste for 4 or 5 weeks and really, really moaning, and I think (.) I know it's horrible, but some days I think, like, I think some people are a lot further down the line than that.* [Emma]

Conversely, several participants recognized the benefits of success stories in the groups in providing "hope that there's light at the end of the tunnel" [Emily]. However, those who recovered often left the online peer support groups. Participants proposed possible reasoning, speculating that "it could trigger some like feelings and memories" [Oliver]. Therefore, success stories were not considered heavily prevalent.

XSL•FO
RenderX

Indeed, when discussions arose surrounding comparing severity of suffering against others in the groups, James noted he found it challenging "just reading the posts because it-it brings up memories of kind of when I was (.) more unwell." Additionally, some participants reported that seeing reports of worse suffering elicited specific feelings:

> *[…] a part of me feels a little bit guilty when I read that, and I think oh (chuckles), why am I moaning about my kind of, minor symptoms.* [Natalie]

Therefore, where participants were in their recovery journeys impacted their experience in the groups.

Administration and Moderation

Having peer support groups on relatively open platforms, such as Facebook, unsurprisingly led to discussion around access and moderation. Many groups utilized gatekeeping questions to allow access. Some participants valued this relative privacy. However, James, Oliver, and Mia all expressed concerns; people could simply lie to get into groups. Therefore, this added (albeit imperfect) security held an important purpose in safeguarding users who reasonably believed they were sharing concerns among peers rather than the general public.

Additionally, immense burdens were placed on administrators and moderators for these support groups:

> *Especially if it's just like 1 person that's just decided to make a group that's all of a sudden got 5000 people with 2000 posts a day.* [James]

Sophia believed responsibility ought to be shared across group members, as "the admins are also people suffering with long COVID," to avoid overburdening these few individuals.

Theme 5: Social Connections

Finding Others

Every participant reported joining groups to feel less alone, with several stating, "I was the only kind of person that I knew around me who'd had COVID" [Emily]. There was appreciation for social media's ability to facilitate connection:

> *[…] thank God we've got all of this online stuff - […] Because if this paned-pandemic had happened when I was a child or whatever, we wouldn't have any of it!* [Claire]

However, James' experience provided an intriguing caveat to the expectation that these platforms can inherently improve connectedness:

> *Just feeling pretty isolated with it all, and okay, there are other people going through this, but (..) (sigh) That - again it's (.) that anonymity of different people it (.) it doesn't really feel - although there is a-a (.) some camaraderie in there, it does kind of almost have that negative impact of feeling kind of more alone (chuckles), conversely.*

Indeed, this increased connection with strangers on the internet was not always perceived as a benefit but rather a notable risk.

> *[…] people are able to get so much detail about you as a person, so sometimes, that does stop me on commenting on things […].* [Mia]

> *[…] if someone wants to talk to me about COVID, they can talk to me about it on the group […] I would prefer to do that in an open forum, where there are admins and if, you know, someone does step over the line, I can sort of say, whoa, too far. So I just - I just delete any private request messages that come through. Um, but I think […] there are a lot of vulnerable people in those groups that (.) that may not do that, and therefore then could become a target of various different things.* [Sophia]

Numerous participants wished for different support group structures to better and more naturally facilitate social connection. Many wanted in-person support, saying that "it would be nice just to sit in a room with 3 or 4 other people and chat through stuff" [James], and Emily noted that this structure could facilitate socializing after periods of isolation.

In the absence of in-person options, either owing to COVID-19 risk or geographical variation among those needing support, several participants noted potential benefits of smaller Zoom calls:

> *[…] it's nicer to interact with someone with say face to face than it is over a keyboard, isn't it?* [Will]

Importantly, some participants wanted smaller groups (either in their current format or as face to face). However, Sophia did not want to lose the unique benefits that larger and more diverse groups provide:

> *[…] some of the articles that I've found most useful are articles that people have posted because they've got completely different issues to me.*

Reassurance

In finding others, participants felt validated and reassured that they were not alone. Several participants used similar language reflecting fears that their symptoms may be psychosomatic. These groups provided spaces to reassure them that these fears were unfounded.

> *[…] it was nice to know that people were going through exactly the same thing, that it was wasn't (…) almost (.) in my head.* [Jessica]

> *But when you see huge volumes of other people reporting the same kind of symptoms, […] it reinforces, and you think well actually, this isn't something that I just made up, […] I'm not being a hypochondriac or, losing my mind.* [Natalie]

Impact on Mental Well-Being

Overall, participants felt these groups helped support their mental well-being in the absence of other care: as encouragement, validation, or an outlet. Chloe's description of the groups was particularly salient:

> *[…] these have been a lifeline for so many people, because (.) when the medical services were failing, this was a beacon of light for people.*

XSL•FO
RenderX

Participants reported that groups provided support and "would lift" them [Chloe], and they were spaces for unpacking challenging emotions among peers who could better understand.

> *I just had to turn down my dream job because I'm not well enough to do it. And I just (.)- I was at the bottom of a pit. […] And I just needed to wallow. And that sort of - it gave me somewhere that I could wallow, even though I always try to be positive […]- I just need a few days to wallow in self-pity, and then I'll - then I'll be alright again.* [Sophia]

However, for many participants, these groups negatively impacted their sense of mental well-being and forced them to navigate the groups with particular attention to the emotional toll.

> *[…] sometimes it can be a little bit overwhelming. I feel sometimes a little bit bombarded with the amount of, sort of, everyone sharing their, sort of, outpourings, of you know. I've literally had people saying they feel suicidal, and - […] You know, that's quite tough to-to read.* [Natalie]

> *Sounds a bit selfish really, but you know, it's nice to check in when you need it really, and to contribute when you need to. Um and then step away from.* [Mia]

Evidently, these peer support groups imperfectly filled gaps out of necessity. Natalie posed an insightful question on concerns around mental health provision more broadly:

> *[…] is that filling a void where there should be more mental health support being offered? And, as we know, particularly in this country, there is no mental health support on the NHS, it's pretty much nonexistent.* [Natalie]

Similarly, several participants expressed how peer support had been, or they wish it had been, in past experiences to help improve their mental well-being. Natalie shared how she was left searching for support after a terrorist attack:

> *I knew other people were struggling and other people were then being impacted. […] And I think at times it would've been really nice to have been able to talk about the experience with them? And kind of, share it, and then - obviously not to dwell on it too much, but just share it, kind of deal with it, and then, be able to move on […].*

When utilized appropriately, peer support could provide immense value in helping people heal and move forward, while supporting their mental well-being.

Discussion

Filling the Gaps: What Led People to Groups?

The results evidence how online peer support groups took on vast roles, exemplified in Figure 1, significantly impacting users. Largely, why participants turned to these groups reflected the existing literature, particularly for contested conditions where dismissal is common [9,56]. Participants focused on how online peer support groups filled health care gaps, frequently reporting how HCPs dismissed concerns and disregarded long COVID lived realities. These groups then allowed for condition validation and formed a sense of solidarity against a medical field that is providing insufficient support, as Barker identified with online groups for fibromyalgia [21].

Furthermore, participants' attempts to seek support reflected experiences by patients with long COVID outside this study, often facing this "postcode lottery" [9,30]. Even when those with long COVID received care, many encountered insufficient compassion or low quality of care [9,12]. James' experience of being "sat in the middle" is all too common. Participants were left with little choice but to seek out support groups, stumbling upon those on Facebook. Equally, participants reported that they struggled to talk about long COVID with loved ones who could (or would) not understand: a common issue with contested conditions [57-59]. As Allen et al [58] noted with online support for long-term conditions, participants here joined these groups to directly compensate for various unmet needs, whether they be medical, informational, or emotional. The unprompted language of feeling "in the dark" that Natalie and Oliver both used was particularly noteworthy, emphasizing these vast knowledge gaps that online peer support groups had to fill.

Complexity and Tensions: Unpacking Varying Experiences

Participants used online peer support groups for social connection through shared experience, as in the literature [35,60]. A notable dichotomy was some used groups if they were feeling low, while others avoided them in these moments. Deciding when to engage elucidates participants' different expectations of perceived benefit. Similarly, engagement was fluid: Participants felt fewer benefits over time, their access to external support changed (for better or worse), or their symptoms fluctuated. Therefore, the flexibility afforded by online peer support groups was distinctly advantageous in providing control [35,61]. Additionally, participants felt empowered when they could support others, as upheld by the literature [62,63]. Of course, individuals in this scenario could face undue burdens [62], though participants did not indicate this, perhaps as responsibility was shared across many group members.

Online support groups provided reassurance. Many participants, like Jessica and Natalie, feared symptoms were all in their head, and thus searched for validation in online peer support groups. Similar to Ziebland and Wyke [35], participants felt a sense of support in these groups, particularly in proving they were not the only ones experiencing this challenging and complex constellation of symptoms. Sharing similar symptomology allowed participants to make sense of their experiences [9,35,59]. Patients coming together around shared illnesses reflects Rabinow's concept of "biosociality" [64] heavily influenced by Foucault's "biopolitics," where bodies are governed through quantification and management [65]. In sharing experiences and knowledge, these groups created "biosocial communities" [66]. Within these communities, "identity work" can occur [59]: members discuss and navigate complex interplays between their condition, society, and self, which brings profound psychological benefit [59]. Unsurprisingly, peer support groups provided space for

XSL•FO
RenderX

emotional support [32,35,56,59,67]. Chloe noted how even just having people write supportive comments profoundly benefited her emotional well-being.

However, in addition to emotional support, participants reported emotional impacts driven by the diversity of long COVID experiences encapsulated in groups, including hope, fear, guilt, and frustration. Participants' experiences reflected Mazanderani et al's [68] concept of being "differently the same" to negotiate shared illness experience. Participants described their relationship to others in online peer support groups as both beneficial and harmful, reflecting Festinger's "social comparison theory" in setting up distinctions against others [69]. In this theory, Festinger [69] noted that people tend to use comparisons to others as a means of better conceptualizing their own situations. In the context of these peer support groups, participants engaged in these social comparisons in various ways. Several participants expressed hope when seeing success stories, an example of upward comparison that helped them visualize the possibility of improvement and thus potentially increased their desire to improve themselves to reach that same goal [70,71]. However, others noted the limited positive outcomes shared in groups, potentially causing fear of sufferings' inevitability and undue stress [72]. Notably, however, upward comparison could cause harm, leading to feelings of jealously and negatively impact self-esteem [71,73,74]. Here, participants were frustrated when seeing those with fewer or less severe symptoms express narratives of deep despondence. Although participants noted users were permitted to use groups in this manner, it dismissed varying experiences. Furthermore, there was evidence of downward comparison in these groups, when individuals look to those who are worse off [75]. This form of comparison caused participants to often feel fear of prolonged suffering, particularly when confronted with narratives from people who experienced long COVID for materially longer than themselves. Interestingly, participants also engaged in downward comparison when they perceived their symptoms to be less "severe" than those of others. This comparison elicited feelings of guilt, though it also created a more implicit sense of gratitude that their experience was not as severe as it could have been.

In contrast to setting situations apart from others, Shiner [76] notes that a central benefit of peer support is the relationality of the peer, particularly in regard to their life experiences. However, with this similarity that emerges inherently in support groups based on particular conditions, there is a notable risk of overidentifying with other's experiences. Ziebland and Wyke [35] identified this risk when investigating the impact of sharing illness experience online. Here, Chloe similarly voiced concerns of people joining online peer support groups too early in their journeys, thus becoming trapped by their suffering.

Who Is the Expert? Examining the Patient-Expert Relationship

Participants felt forced into taking control over their recovery rather than relying on "experts." In doing so, they reflected Rose and Novas' [66] use of "biological citizenship." Individuals are expected to take greater responsibility over their health and greater self-management [66]. Participants reported attempts to fill medical gaps through sharing therapies. Some emphasized potential risks of this activity more than others, highlighting varying recovery approaches. Some tried what others with long COVID suggested, as it was the only practical advice available; others were hesitant to try unverified ideas, thus creating a conflict. Interestingly, numerous participants used a similar language to Oliver: they took "everything with a pinch of salt." This repeated language emphasizes how participants in online peer support groups had to determine safety of suggestions themselves, reflecting burdens placed unduly on patients to critically appraise content. This finding was similar to what Ziebland and Wyke [35] had previously identified; there can be little to distinguish the relative trustworthiness of content that is shared in online spheres.

Patients provided unique forms of knowledge based on embodied illness experience [59,77], often afforded greater credibility by other patients [63]. It is worth invoking Kleinman's [78] disease versus illness distinction: disease as the biological mechanisms doctors focus on and illness as the "innately human experience of symptoms and suffering." Long COVID has impacts beyond corporeal disease definitions prioritized by biomedicine; instead, it embodies illness narratives with implications felt heavily at the experiential level. It is therefore unsurprising that these groups attempted to provide both daily functioning and medically oriented support, potentially at a faster rate than would be gained from HCPs [59]. Of course, as many participants noted, there are significant risks with this role: information is no longer authenticated, and significant harm can arise.

Furthermore, participants reported how online peer support groups could be ahead of biomedicine, challenging biomedical and societal assumptions [9,59]. Traditionally, expert authority is not based on patient experiences [35], though participants in these long COVID support groups subverted this narrative. This process is common with contested conditions where patients become "lay experts" [21,59,77,79], unsettling biomedicine's authority [59,79]. Although professionalization is considered beneficial in empowering patients [59], the results from this study unveil a slightly different story. Of course, having the ability to retain control is essential to avoid becoming entirely disempowered by a dismissive health care system. Nevertheless, embodying expert roles should not necessarily be strived for. Many participants expressed frustration at how people were forced into this role. Despite lay expertise's benefits, there is continuing conflict between professional and lay experts. Several participants reported taking information from groups to HCPs to guide discussion and advocate for care: a dynamic previously identified in giving patients greater control over care [35,59,80]. Indeed, this behavior could help patients avoid unnecessary tests and appointments [35]. However, HCPs can react negatively if patients' ideas conflict with their recommendations [81]. Additionally, HCPs remain gatekeepers to care [21]. As a result, interventions available to participants were often limited to over-the-counter medications and at-home exercises.

Improving the Groups: Participant's Desires

The literature indicates that online peer support groups offer comparable support to in-person support, with distinct

XSL•FO
RenderX

advantages (eg, asynchronous and greater anonymity) [35,61]. Conversely, participants in this study expressed that face-to-face peer support would enhance the experience, facilitating greater connectivity and emotional support without restricted communication "over a keyboard" [Will]. This real-world connection could avoid a potential pitfall of online peer support groups that Ziebland and Wyke [35] identified, where users could get too absorbed in the virtual world at the detriment of their external social worlds. Additionally, participants wished for smaller groups, struggling to feel connected to others in large global groups. Jessica's overwhelmingly positive experience in her small WhatsApp group could reflect what other participants desired: a closer-knit community with naturally fostered social connection. The craving for more in-person connection may be partially due to reduced face-to-face connection throughout COVID-19 lockdowns; as Emily stated, these groups could aid transitions back into socializing.

Furthermore, the responsibility placed (as participants noted, unduly) on those experiencing long COVID brought up questions about where responsibility ought to lie. With self-management, peers could inappropriately be considered replacements to medical services [32]. Jessica's experience is an interesting point of comparison. In her case, doctors set up the WhatsApp group and then left, allowing participants to engage without oversight. She appreciated this setup greatly. Those in the Facebook online peer support groups often wished for more professional involvement, either in setup or in content moderation. Having a professional moderator is atypical in online peer support groups, but it can promote engagement [60]. Participants made it clear that there were notable flaws alongside the benefits in the current structure and content of these groups. However, Natalie's commentary on peer support in other contexts highlights how, if used appropriately, these groups could have immense benefit for those with long COVID.

Limitations

The relatively small sample size that I used owing to the practical constraints discussed previously had implications on the study. Results may not fully capture the range of experiences that individuals may have in these groups, though it became clear after the first few interviews that consensus was emerging around possible themes. Additionally, the smaller sample size allowed for rich qualitative data generation, providing valuable insights as research begins to explore these online peer support groups in long COVID recovery.

Additionally, selection bias is a risk as individuals volunteered to participate: volunteers' experiences may not be representative of others. In addition, in sharing my recruitment poster in larger groups to increase outreach, I missed smaller groups. In using a first-come, first-served recruitment methodology due to significant time limitations within my master's timeline, diversity of participants was limited to those who responded promptly to my recruitment poster. Consequently, generalizability to others in all long COVID online peer support groups is limited.

Furthermore, Zoom interviews had limitations. Individuals with lower technology literacy or limited access to required technology may have been excluded [40]. In addition, 6 interviews experienced technical difficulties. However, I used these technical challenges to build rapport, easing tensions, breaking down power hierarchies, and promptly problem-solving together [41].

Implications

These results fit into the wider discourse, with implications for global public health policy, practice, and research. This study could encourage improvements in the United Kingdom's long COVID programs to reflect patient needs rather than perceived needs from policy makers and HCPs: embedding appropriate peer support within broader and accessible medically oriented care. Although this study's UK focus permitted more directed analysis, many groups were international, suggesting possible universality of long COVID care shortfalls. Insights here could aid stakeholders globally in designing, implementing, or participating in long COVID care and peer support specifically. However, this study's findings may not be entirely transferrable to different health care and sociocultural contexts. For example, in countries with health care systems requiring out-of-pocket payments or complex insurance policies, the reasons to use and the importance of these peer support groups may differ without the same "safety net" of the United Kingdom's universal health care system. Further research from other countries is essential.

This paper provides an overview of experiences in long COVID online peer support groups. Future research could delve deeper into each subtheme or could explore these groups' roles and importance among those often marginalized in health care provision (including ethnic minorities, lower socioeconomic groups, and those with disabilities).

Conclusion

Online peer support groups were *a lifeline but insufficient*. They were imperfect but were needed to fill immense gaps in health care and social support. This study fills epistemological gaps on lived experience of long COVID: beyond corporeal suffering into ways people navigate their newfound reality with a medical field that has yet to catch up. As more attention is given to the condition, hopefully the dark clouds obscuring long COVID will begin to lift, paving the way for more attuned and appropriate care.

Acknowledgments

The author wishes to express thanks to the faculty at the London School of Hygiene & Tropical Medicine for their encouragement throughout the research process. Additionally, the author expresses immense gratitude to the participants of this study for their willingness to be vulnerable and open. The author also wishes to thank Jessica Amick for their time and insights in the peer-debriefing process.

XSL•FO
RenderX

Data Availability

The data generated or analyzed during the study are available from the corresponding author on reasonable request. Available data are in the form of a sufficiently anonymized Microsoft Excel spreadsheet containing all the coded extracts organized into themes and subthemes during thematic analysis. The raw data (transcripts) are not available, as participants provided consent on the basis that only the researcher would have access to these documents.

Authors' Contributions

HD was the sole author and contributor to this research and paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview topic guide.
[DOCX File , 24 KB - jmir_v24i5e37674_app1.docx]

Multimedia Appendix 2
Participant information sheet.
[DOCX File , 674 KB - jmir_v24i5e37674_app2.docx]

Multimedia Appendix 3
Resources for further support.
[DOCX File , 511 KB - jmir_v24i5e37674_app3.docx]

References

1. GOV.UK. Cases in United Kingdom. URL: https://coronavirus.data.gov.uk/details/cases [accessed 2022-03-30]
2. GOV.UK. Deaths in United Kingdom. URL: https://coronavirus.data.gov.uk/details/deaths [accessed 2022-03-30]
3. Harding L. 'Weird as hell': The Covid-19 Patients Who Have Symptoms for Months. 2020 May 15. URL: http://www.theguardian.com/world/2020/may/15/weird-hell-professor-advent-calendar-covid-19-symptoms-paul-garner [accessed 2021-09-01]
4. National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN), Royal College of General Practitioners (RCGP). COVID-19 Rapid Guideline: Managing the Long-Term Effects of COVID-19. 2021. (1.8) URL: https://www.nice.org.uk/guidance/ng188/resources/covid19-rapid-guideline-managing-the-longterm-effects-of-covid19-pdf-51035515742 [accessed 2022-01-09]
5. Office for National Statistics. Prevalence of Ongoing Symptoms Following Coronavirus (COVID-19) Infection in the UK: 7 April 2022. 2022 Apr 7. URL: https://tinyurl.com/3p7ut9u5 [accessed 2022-04-08]
6. Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent symptoms following SARS-CoV-2 infection in a random community sample of 508,707 people. medRxiv Preprint posted online July 3, 2021. [doi: 10.1101/2021.06.28.21259452]
7. Office for National Statistics. Prevalence of Ongoing Symptoms Following Coronavirus (COVID-19) Infection in the UK: 4 June 2021. 2021 Jun 4. URL: https://tinyurl.com/yduhmhby [accessed 2022-01-12]
8. Logue JK, Franko NM, McCulloch DJ, McDonald D, Magedson A, Wolf CR, et al. Sequelae in adults at 6 months after COVID-19 infection. JAMA Netw Open 2021 Feb 19;4(2):e210830 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.0830] [Medline: 33606031]
9. Rushforth A, Ladds E, Wieringa S, Taylor S, Husain L, Greenhalgh T. Long Covid - the illness narratives. Soc Sci Med 2021 Oct;286:114326 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114326] [Medline: 34425522]
10. Cabinet Office. COVID-19 Response: Living with COVID-19. 2022 Feb 21. URL: https://www.gov.uk/government/publications/covid-19-response-living-with-covid-19 [accessed 2022-02-21]
11. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. EClinicalMedicine 2021 Aug 1;38:101019 [FREE Full text] [doi: 10.1016/j.eclinm.2021.101019] [Medline: 34308300]
12. The Lancet. Facing up to long COVID. Lancet 2020 Dec;396(10266):1861. [doi: 10.1016/s0140-6736(20)32662-3]
13. Sudre C, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and predictors of long-COVID: analysis of COVID cases and their symptoms collected by the Covid Symptoms Study app. medRxiv Preprint posted online Dec 19, 2020. [doi: 10.1101/2020.10.19.20214494]
14. Office for National Statistics. Coronavirus and the Social Impacts of 'Long COVID' on People's Lives in Great Britain: 7 April to 13 June 2021. 2020 Jul 21. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/

XSL•FO
RenderX

conditionsanddiseases/articles/coronavirusandthesocialimpactsoflongcovidonpeopleslivesingreatbritain/7aprilto13june2021 [accessed 2021-08-28]
15. Nabavi N. Long covid: how to define it and how to manage it. BMJ 2020 Sep 07;370:m3489. [doi: 10.1136/bmj.m3489] [Medline: 32895219]
16. Callard F, Perego E. How and why patients made long Covid. Soc Sci Med 2021 Jan;268:113426 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113426] [Medline: 33199035]
17. Perego E. @elisaperego78. 2020 May 20. URL: https://twitter.com/elisaperego78/status/1263172084055838721 [accessed 2021-10-02]
18. World Health Organization. A Clinical Case Definition of Post COVID-19 Condition by a Delphi Consensus, 6 October 2021. 2021. URL: https://www.who.int/publications-detail-redirect/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1 [accessed 2022-03-30]
19. Conrad P, Schneider J. Deviance and Medicalization: From Badness to Sickness. Philadelphia, PA: Temple University Press; 1992.
20. Illich I. Medical Nemesis: The Expropriation of Health. New York, NY: Pantheon Books; 1976.
21. Barker KK. Electronic support groups, patient-consumers, and medicalization: the case of contested illness. J Health Soc Behav 2008 Mar 01;49(1):20-36. [doi: 10.1177/002214650804900103] [Medline: 18418983]
22. National Health Service (NHS). Long COVID: The NHS Plan for 2021/22. 2021 Jun 15. URL: https://www.england.nhs.uk/coronavirus/publication/long-covid-the-nhs-plan-for-2021-22/ [accessed 2021-08-28]
23. National Health Service (NHS). Post-COVID Syndrome (Long COVID). URL: https://www.england.nhs.uk/coronavirus/post-covid-syndrome-long-covid/ [accessed 2022-01-16]
24. Bradley J. Scotland Will Not Create “long Covid” Clinics due to Lack of Understanding of Condition. 2021 Jun 10. URL: https://tinyurl.com/4awjtws8 [accessed 2022-01-16]
25. Godden S, McCool M. Thousands “Not Referred” to Long Covid Support Service. 2021 Sep 2. URL: https://www.bbc.com/news/uk-scotland-58409187 [accessed 2022-01-16]
26. Martin P. Long Covid Patients Call for Specialist Clinics in Wales. 2021 Jun 30. URL: https://www.bbc.com/news/uk-wales-57654683 [accessed 2022-01-16]
27. Department of Health. Services for Patients with Post Covid-19 Syndrome (PCS) Launched. 2021 Nov 01. URL: https://www.health-ni.gov.uk/news/services-patients-post-covid-19-syndrome-pcs-launched [accessed 2022-01-16]
28. Lokugamage AU, Taylor S, Rayner C. Patients’ Experiences of “Longcovid” Are Missing from the NHS Narrative. 2020 Jul 10. URL: https://blogs.bmj.com/bmj/2020/07/10/patients-experiences-of-longcovid-are-missing-from-the-nhs-narrative/ [accessed 2021-10-02]
29. Jensen L. "Just stay at home" was lonely and terrifying. BMJ 2020 Nov 05;371:m3807. [doi: 10.1136/bmj.m3807] [Medline: 33154066]
30. Geddes L. UK’s Long Covid Patients Facing Postcode Lottery for Support. 2021 Apr 06. URL: http://www.theguardian.com/society/2021/apr/06/uk-long-covid-patients-facing-postcode-lottery-for-support [accessed 2021-08-28]
31. Gorna R, MacDermott N, Rayner C, O’Hara M, Evans S, Agyen L, et al. Long COVID guidelines need to reflect lived experience. Lancet 2021 Feb;397(10273):455-457. [doi: 10.1016/s0140-6736(20)32705-7]
32. Dennis C. Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003 Mar;40(3):321-332. [doi: 10.1016/S0020-7489(02)00092-5]
33. Abdi F, Simbar M. The peer education approach in adolescents: narrative review article. Iran J Public Health 2013 Nov;42(11):1200-1206 [FREE Full text] [Medline: 26171331]
34. Kim JN, Lee S. Communication and cybercoping: coping with chronic illness through communicative action in online support networks. J Health Commun 2014 Feb;19(7):775-794. [doi: 10.1080/10810730.2013.864724] [Medline: 24559492]
35. Ziebland S, Wyke S. Health and illness in a connected world: how might sharing experiences on the internet affect people's health? Milbank Q 2012 Jun;90(2):219-249 [FREE Full text] [doi: 10.1111/j.1468-0009.2012.00662.x] [Medline: 22709387]
36. Green J, Thorogood N. Qualitative Methods for Health Research. New York, NY: SAGE Publications; 2018.
37. Dreyer K, Christopher RS. An Introduction to Phenomenological Psychology. Pittsburgh, PA: Duquesne University Press; 1981.
38. Kvale S. InterViews: An Introduction to Qualitative Research. New York, NY: SAGE Publications; 1996.
39. Eysenbach G, Wyatt J. Using the internet for surveys and health research. J Med Internet Res 2002 Nov;4(2):E13 [FREE Full text] [doi: 10.2196/jmir.4.2.e13] [Medline: 12554560]
40. Hensen B, Mackworth-Young CRS, Simwinga M, Abdelmagid N, Banda J, Mavodza C, et al. Remote data collection for public health research in a COVID-19 era: ethical implications, challenges and opportunities. Health Policy Plan 2021 Apr 21;36(3):360-368 [FREE Full text] [doi: 10.1093/heapol/czaa158] [Medline: 33881138]
41. Archibald MM, Ambagtsheer RC, Casey MG, Lawless M. Using Zoom videoconferencing for qualitative data collection: perceptions and experiences of researchers and participants. Int J Qual Methods 2019 Sep 11;18:160940691987459. [doi: 10.1177/1609406919874596]
42. Poland BD. Transcription quality. In: Holstein JA, Gubrium JF, editors. Inside Interviewing. New Lenses, New Concerns. New York, NY: SAGE Publications; 2003:568.

XSL•FO
RenderX

43. Merriam S. Qualitative Research: A Guide to Design and Implementation. New York, NY: John Wiley & Sons; 2009.
44. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
45. Lincoln YS, Guba EG. Naturalistic Inquiry. New York, NY: SAGE Publications; 1985:308.
46. Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory Pract 2000 Aug;39(3):124-130. [doi: 10.1207/s15430421tip3903_2]
47. Spall S. Peer debriefing in qualitative research: emerging operational models. Qual Inq 1998 Jul;4(2):280-292. [doi: 10.1177/107780049800400208]
48. Peterman A, Bhatia A, Guedes A. Remote Data Collection on Violence against Women during COVID-19: A Conversation with Experts on Ethics, Measurement & Research Priorities (Part 1). 2020 May 11. URL: https://www.unicef-irc.org/article/1997-remote-data-collection-on-violence-against-women-during-covid-19-a-conversation-with.html [accessed 2022-01-29]
49. Oates J. BPS Code of Human Research Ethics. Leicester, UK: British Psychological Society; 2021.
50. Reid AM, Brown JM, Smith JM, Cope AC, Jamieson S. Ethical dilemmas and reflexivity in qualitative research. Perspect Med Educ 2018 Mar 13;7(2):69-75 [FREE Full text] [doi: 10.1007/s40037-018-0412-2] [Medline: 29536374]
51. Ellingson LL. Embodied knowledge: writing researchers' bodies into qualitative health research. Qual Health Res 2006 Feb 01;16(2):298-310. [doi: 10.1177/1049732305281944] [Medline: 16394216]
52. Brannick T, Coghlan D. In defense of being "native": the case for insider academic research. Organ Res Methods 2007 Jan 1;10(1):59-74. [doi: 10.1177/1094428106289253]
53. Mercer J. The challenges of insider research in educational institutions: wielding a double‐edged sword and resolving delicate dilemmas. Oxford Rev Educ 2007 Feb;33(1):1-17. [doi: 10.1080/03054980601094651]
54. Chavez C. Conceptualizing from the inside: advantages, complications, and demands on insider positionality. Qual Rep 2008 Aug 14;13:474-494. [doi: 10.46743/2160-3715/2008.1589]
55. Harris M. "Three in the room": embodiment, disclosure, and vulnerability in qualitative research. Qual Health Res 2015 Dec 09;25(12):1689-1699. [doi: 10.1177/1049732314566324] [Medline: 25576481]
56. Dumit J. Illnesses you have to fight to get: facts as forces in uncertain, emergent illnesses. Soc Sci Med 2006 Feb;62(3):577-590. [doi: 10.1016/j.socscimed.2005.06.018] [Medline: 16085344]
57. Armentor JL. Living with a contested, stigmatized illness: experiences of managing relationships among women with fibromyalgia. Qual Health Res 2017 Mar 10;27(4):462-473. [doi: 10.1177/1049732315620160] [Medline: 26667880]
58. Allen C, Vassilev I, Kennedy A, Rogers A. Long-term condition self-management support in online communities: a meta-synthesis of qualitative papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
59. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan 12;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
60. Maclachlan LR, Mills K, Lawford BJ, Egerton T, Setchell J, Hall LM, et al. Design, delivery, maintenance, and outcomes of peer-to-peer online support groups for people with chronic musculoskeletal disorders: systematic review. J Med Internet Res 2020 Apr 24;22(4):e15822 [FREE Full text] [doi: 10.2196/15822] [Medline: 32329746]
61. Brunelli AA, Murphy GC, Athanasou JA. Effectiveness of social support group interventions for psychosocial outcomes: a meta-analytic review. Aust J Rehabil Couns 2016 Dec 13;22(2):104-127. [doi: 10.1017/jrc.2016.9]
62. Bracke P, Christiaens W, Verhaeghe M. Self-esteem, self-efficacy, and the balance of peer support among persons with chronic mental health problems. J Appl Soc Psychol 2008;38(2):436-459. [doi: 10.1111/j.1559-1816.2008.00312.x]
63. Aboumatar H, Naqibuddin M, Neiman J, Saunders J, Kim S, Chaudhry H, et al. Methodology and baseline characteristics of a randomized controlled trial testing a health care professional and peer-support program for patients with chronic obstructive pulmonary disease: the BREATHE2 study. Contemp Clin Trials 2020 Jul;94:106023. [doi: 10.1016/j.cct.2020.106023] [Medline: 32360887]
64. Rabinow P. Artificiality and enlightenment: from sociobiology to biosociality. In: Essays on the Anthropology of Reason. Princeton, NJ: Princeton University Press; 1996:99-111.
65. Foucault M. The birth of biopolitics. In: Rabinow P, editor. Ethics, Subjectivity and Truth. The Essential Works of Michel Foucault, 1954-1984. Vol 1. New York, NY: New Press; 1997:73-79.
66. Rose N, Novas C. Biological ctizenship. In: Ong A, Collier SJ, editors. Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems. Hoboken, NJ: Blackwell; 2004:439-463.
67. Milburn K. A critical review of peer education with young people with special reference to sexual health. Health Educ Res 1995 Dec;10(4):407-420. [doi: 10.1093/her/10.4.407] [Medline: 10159674]
68. Mazanderani F, Locock L, Powell J. Being differently the same: the mediation of identity tensions in the sharing of illness experiences. Soc Sci Med 2012 Feb;74(4):546-553. [doi: 10.1016/j.socscimed.2011.10.036] [Medline: 22227237]
69. Festinger L. A theory of social comparison processes. Hum Relat 1954 May;7(2):117-140. [doi: 10.1177/001872675400700202]
70. Wheeler L. Motivation as a determinant of upward comparison. J Exp Soc Psychol 1966 Sep;1:27-31. [doi: 10.1016/0022-1031(66)90062-x]

XSL•FO
RenderX

71. Collins RL. For better or worse: the impact of upward social comparison on self-evaluations. Psychol Bull 1996;119(1):51-69 [FREE Full text] [doi: 10.1037/0033-2909.119.1.51]
72. Salzer MS, Palmer SC, Kaplan K, Brusilovskiy E, Ten Have T, Hampshire M, et al. A randomized, controlled study of Internet peer-to-peer interactions among women newly diagnosed with breast cancer. Psychooncology 2010 Apr;19(4):441-446. [doi: 10.1002/pon.1586] [Medline: 19484712]
73. Wang JL, Wang HZ, Gaskin J, Hawk S. The mediating roles of upward social comparison and self-esteem and the moderating role of social comparison orientation in the association between social networking site usage and subjective well-being. Front Psychol 2017 May 11;8:771 [FREE Full text] [doi: 10.3389/fpsyg.2017.00771] [Medline: 28553256]
74. Krasnova H, Widjaja T, Buxmann P, Wenninger H, Benbasat I. Why following friends can hurt you: an exploratory investigation of the effects of envy on social networking sites among college-age users. Inf Syst Res 2015 Sep;26(3):585-605. [doi: 10.1287/isre.2015.0588]
75. Wills TA. Downward comparison principles in social psychology. Psychol Bull 1981;90(2):245-271 [FREE Full text] [doi: 10.1037/0033-2909.90.2.245]
76. Shiner M. Defining peer education. J Adolesc 1999 Aug;22(4):555-566. [doi: 10.1006/jado.1999.0248] [Medline: 10469518]
77. Hartzler A, Pratt W. Managing the personal side of health: how patient expertise differs from the expertise of clinicians. J Med Internet Res 2011 Aug 16;13(3):e62 [FREE Full text] [doi: 10.2196/jmir.1728] [Medline: 21846635]
78. Kleinman A. The Illness Narratives: Suffering, Healing, and the Human Condition. New York, NY: Basic Books; 1998.
79. Kroll-Smith S, Floyd HH. Bodies in Protest: Environmental Illness and the Struggle over Medical Knowledge. New York, NY: New York University Press; 1997.
80. McMullan M. Patients using the internet to obtain health information: how this affects the patient-health professional relationship. Patient Educ Couns 2006 Oct;63(1-2):24-28. [doi: 10.1016/j.pec.2005.10.006] [Medline: 16406474]
81. Francis H, Carryer J, Wilkinson J. Patient expertise: contested territory in the realm of long-term condition care. Chronic Illn 2019 Sep 22;15(3):197-209. [doi: 10.1177/1742395318757853] [Medline: 29466874]

Abbreviations

HCP: health care professional
NHS: National Health Service
NICE: National Institute for Health and Care Excellence

Edited by G Eysenbach; submitted 02.03.22; peer-reviewed by M Wasilewski, L Locock; comments to author 19.03.22; revised version received 17.04.22; accepted 20.04.22; published 20.05.22.

Please cite as:
Day HLS
Exploring Online Peer Support Groups for Adults Experiencing Long COVID in the United Kingdom: Qualitative Interview Study
J Med Internet Res 2022;24(5):e37674
URL: https://www.jmir.org/2022/5/e37674
doi:10.2196/37674
PMID:35468083

©Hannah L S Day. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 20.05.2022. 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.

XSL•FO
RenderX

Original Paper

Integration of an Intensive Care Unit Visualization Dashboard (i-Dashboard) as a Platform to Facilitate Multidisciplinary Rounds: Cluster-Randomized Controlled Trial

Chao-Han Lai[1,2,3], MD, PhD; Kai-Wen Li[4], BSc; Fang-Wen Hu[4], PhD; Pei-Fang Su[5], PhD; I-Lin Hsu[1], MD, PhD; Min-Hsin Huang[1], MD, MSc; Yen-Ta Huang[1], MD, PhD; Ping-Yen Liu[6,7,8], MD, PhD; Meng-Ru Shen[9,10], MD, PhD

[1]Department of Surgery, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[2]Department of Biochemistry and Molecular Biology, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

[4]Department of Nursing, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[5]Department of Statistics, College of Management, National Cheng Kung University, Tainan City, Taiwan

[6]Division of Cardiology, Department of Internal Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[7]Institute of Clinical Medicine, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[8]Department of Clinical Medical Research, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[9]Department of Obstetrics and Gynecology, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan

[10]Department of Pharmacology, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

Corresponding Author:
Meng-Ru Shen, MD, PhD
Department of Obstetrics and Gynecology
National Cheng Kung University Hospital, College of Medicine
National Cheng Kung University
138, Sheng-Li Road
Tainan, 704
Taiwan
Phone: 886 062353535 ext 5505
Fax: 886 2766676
Email: mrshen@mail.ncku.edu.tw

Abstract

Background: Multidisciplinary rounds (MDRs) are scheduled, patient-focused communication mechanisms among multidisciplinary providers in the intensive care unit (ICU).

Objective: *i*-Dashboard is a custom-developed visualization dashboard that supports (1) key information retrieval and reorganization, (2) time-series data, and (3) display on large touch screens during MDRs. This study aimed to evaluate the performance, including the efficiency of prerounding data gathering, communication accuracy, and information exchange, and clinical satisfaction of integrating *i*-Dashboard as a platform to facilitate MDRs.

Methods: A cluster-randomized controlled trial was performed in 2 surgical ICUs at a university hospital. Study participants included all multidisciplinary care team members. The performance and clinical satisfaction of *i*-Dashboard during MDRs were compared with those of the established electronic medical record (EMR) through direct observation and questionnaire surveys.

Results: Between April 26 and July 18, 2021, a total of 78 and 91 MDRs were performed with the established EMR and *i*-Dashboard, respectively. For prerounding data gathering, the median time was 10.4 (IQR 9.1-11.8) and 4.6 (IQR 3.5-5.8) minutes using the established EMR and *i*-Dashboard (P<.001), respectively. During MDRs, data misrepresentations were significantly less frequent with *i*-Dashboard (median 0, IQR 0-0) than with the established EMR (4, IQR 3-5; P<.001). Further, effective recommendations were significantly more frequent with *i*-Dashboard than with the established EMR (P<.001). The questionnaire results revealed that participants favored using *i*-Dashboard in association with the enhancement of care plan development and team participation during MDRs.

XSL•FO
RenderX

Conclusions: *i*-Dashboard increases efficiency in data gathering. Displaying *i*-Dashboard on large touch screens in MDRs may enhance communication accuracy, information exchange, and clinical satisfaction. The design concepts of *i*-Dashboard may help develop visualization dashboards that are more applicable for ICU MDRs.

Trial Registration: ClinicalTrials.gov NCT04845698; https://clinicaltrials.gov/ct2/show/NCT04845698

(J Med Internet Res 2022;24(5):e35981) doi:10.2196/35981

KEYWORDS

Intensive care unit; multidisciplinary round; visualization dashboard; large screen; information management strategy; electronic health record; medical record; digital health; dashboard; i-Dashboard; electronic medical record; information exchange

Introduction

Medical care in intensive care units (ICUs) consumes a substantial part of the income of many countries worldwide, and the enormous burden continues to grow [1,2]. Integrated multidisciplinary teamwork, a patient-centered model of care in which intensivists and other members from relevant disciplines provide critical care as a team, effectively complements intensivist care and improves outcomes for critically ill medical and surgical patients [3,4]. Multidisciplinary rounds (MDRs; also called *interprofessional rounds*) are mechanisms that involve scheduled discussion among multidisciplinary providers, including physicians, registered nurses, nurse practitioners (NPs), respiratory therapists (RTs), pharmacists, and dietitians, to review clinical information, exchange opinions, and develop a plan of care [5]. Because effective communication among providers is essential to high-quality patient care, failures during this process may potentially impact the safety and outcomes of ICU patients [5-7].

Understanding causes that potentially impede interdisciplinary communication during MDRs may facilitate improvement in the communication quality among multidisciplinary providers. Based on a systematic review of evidence-informed practices for ICU MDRs, poor retrieval of patient information has been identified as a barrier that hinders information exchange [5]. Currently, clinicians manually access patient information from disparate modules in information systems, and data aggregated into electronic medical record (EMR)-generated printouts or handwritten notes are verbalized in MDRs [8,9]. A recent study revealed that nearly 40% of verbalized laboratory data are inaccurately communicated during MDRs, and only 7.8% of data misrepresentations that precipitate erroneous clinical decisions can be detected [8].

One of the objectives of the technological advancements applied to critical care is simplifying all the avenues of information [10]. It appears that visualization dashboards (also called EMR viewers) have great potential to be the solution as these dashboards are known for the efficiency of clinical information management [9,11-13]. Notably, compared to the standard EMR environment, introducing visualization dashboards may not improve perceived satisfaction with MDRs, such as information presentation or team participation and communication [12]. A possible problem is that displaying dashboards on small monitors positioned on a trolley or bedside computers may give unequal access to data and cause a body orientation shift of providers from other participants to monitors [14], thus potentially hampering interdisciplinary communication during MDRs. In addition, because of unequal EMR access for real-time data viewing to recognize errors and the inability to simultaneously listen, process, and verify data, the multidisciplinary care team relies disproportionately on the intensivist to detect data misrepresentations that potentially lead to medical errors [8]. In the era of rapid development of information technology, an integrated information management strategy to facilitate information retrieval and enhance interdisciplinary communication in MDRs remains to be explored.

In this study, we aimed to develop a user experience–oriented platform as an integrated solution to assist MDRs. The *i*-Dashboard is a care team–designed, patient-centered visualization dashboard in which information extracted from different sources was reorganized on the basis of the requirement of different disciplines or transformed into time-series data as needed. During MDRs, *i*-Dashboard is displayed on wall-mounted large touch screens to bring effective visualization to the multidisciplinary care team. We assumed that *i*-Dashboard might aid prerounding data gathering, and integrating *i*-Dashboard displayed on large touchscreens during MDRs might enhance interdisciplinary communication. Thus, the efficiency, communication accuracy, information exchange, and clinical satisfaction of integrating *i*-Dashboard as a platform to facilitate MDRs were evaluated.

Methods

Design and Participants

The study was conducted in a 1300-bed university hospital that offers first-line and tertiary referral services for a population of approximately 1.8 million individuals in southern Taiwan. A cluster-randomized controlled trial was conducted in 2 of the 4 surgical ICUs with 10 and 8 beds. The established EMR (control) and *i*-Dashboard (intervention) were randomly assigned as tools to facilitate prerounding information collection and MDRs in the 2 units and exchanged at 2-week intervals.

Before this trial, MDRs have been carried out in the study units for ~5 years. An integrated multidisciplinary care team is composed of at least 1 intensivist, a registered nurse, an NP, an RT, a pharmacist, and a dietitian. All these providers attend MDRs held on a regular schedule 3 times a week. MDRs are conducted only for patients who stay for more than 7 days because patients receiving surgical ICU care for more than 7 days have a high rate of in-hospital mortality [15], and ~75% of patients have a length of stay for 7 days or less in study units.

XSL•FO
RenderX

Ethical Considerations

This study was registered with ClinicalTrials.gov (NCT04845698). The study protocol was approved by the institutional review board of National Cheng Kung University Hospital (B-ER-110-040). Informed consent was obtained from all participants in the study. This study is reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile HEalth Applications and onLine TeleHealth) checklist (Multimedia Appendix 1) [16].

Established EMR

The established EMR environment applied in the ICU includes the Philips IntelliSpace Critical Care & Anesthesia information system (ICCA; Philips) and the Hospital Information System (HIS) developed by the institutional Department of Information Technology and its subsystems, including the Laboratory Information System (LIS) and the Picture Archiving and Communication System (PACS). Philips ICCA is an ICU-specific EMR system that provides information essential for critical care [17]. Patient data are organized by data category (demographics, vital signs, laboratory data, etc) in a series of tabs or window panels. The HIS supports text data key-in and patient order entry.

Development and Architecture of i-Dashboard

This study evaluated the performance of the first version of *i*-Dashboard. The *i*-Dashboard (Advantech) was custom-developed under the guidance of multidisciplinary professionals, including physicians, registered nurses, NPs, RTs, pharmacists, and dietitians, rather than only physicians because information for MDRs needs may vary on the basis of the clinical role. Every health care provider working in the surgical ICU participated in developing *i*-Dashboard. Different professionals held preparatory meetings of their own. Before *i*-Dashboard was formally implemented for clinical use, the structure and layout were repeatedly revised and tailored to achieve broad acceptance among directors of multidisciplinary providers in the ICU.

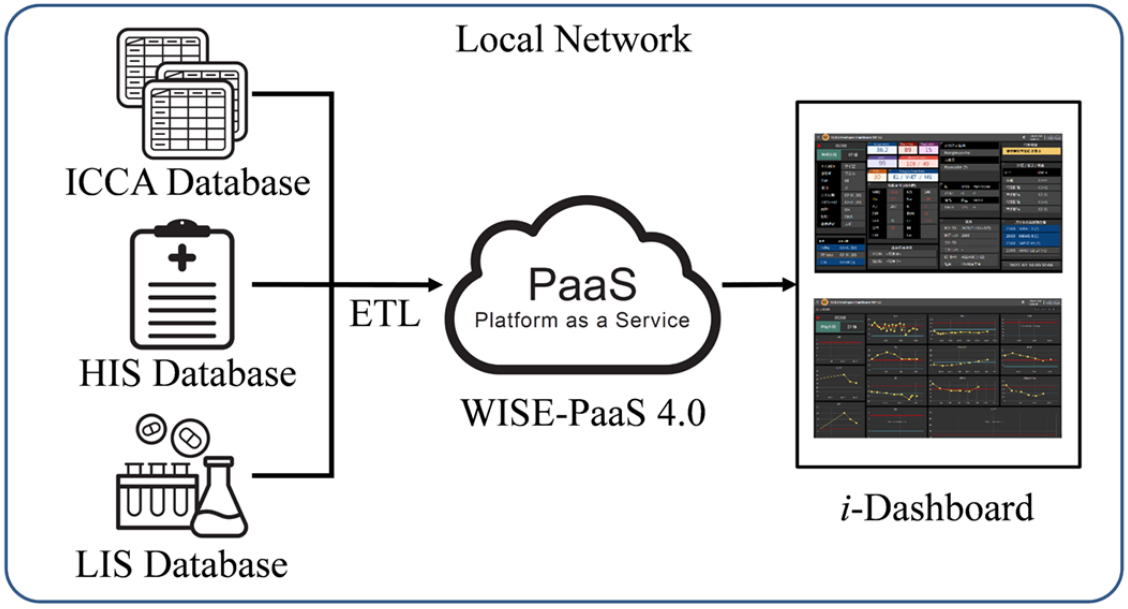
The architecture of *i*-Dashboard is summarized in Multimedia Appendix 2. In the static mode, the *i*-Dashboard was designed to substitute as a station whiteboard with lists of patients, information to aid emergency evacuation, and on-duty physicians, nurses, and NPs. Patient-level data in *i*-Dashboard were modified from the MDR checklist and digitally transformed. Therefore, data were preidentified and retrieved from different origins in the established EMR environment, especially ICCA. Instead of database-centered displays, these data were reorganized on the basis of the requirements of different professionals or disciplines to form dashboard pages (ie, an overview page and an RT-pharmacist-dietitian page) and element blocks. In addition to colored signaling for values outside the reference ranges, *i*-Dashboard was designed to support built-in automated calculation of severity scores (eg, Simplified Acute Physiology Score II and Sequential Organ Failure Assessment) and visualization of time-series data to expedite navigation of patient condition. Time-series data (eg, vital signs, laboratory data, or severity scores) that were transformed into line charts can be accessed through the hyperlinks located at the left upper corner of element blocks on the overview page. The pages of time-series data were designed to support both fixed (eg, last 24 hours or last 3 days) and relative custom time frames are available.

The technical details underlying *i*-Dashboard are summarized in Figure 1. We used a Windows 10 PC as the visualization platform to support *i*-dashboard. A K8s-based WISE-PaaS 4.0 platform (Advantech) facilitates the integration of diverse devices and communication protocols, making data exchange and system development agile. The entire platform was developed and deployed on 6 VMware servers, each with a 256-GB hard drive and 32-GB memory, and a 24-core Intel Xeon Gold 6248R 3.00 GHz processor. The servers received the data from the Philips ICCA, HIS, and LIS database servers.

In the study units, both the established EMR and *i*-Dashboard can be accessed through desktop computers with 17- or 22-inch monitors or mobile platforms.

XSL•FO
RenderX

Figure 1. Transfer of the IntelliSpace Critical Care & Anesthesia information system (ICCA), Hospital Information System (HIS), and Laboratory Information System (LIS) data to i-Dashboard. ETL: Extract-Transform-Load. PaaS: Platform as a Service. WISE-PaaS 4.0: brand name of the platform belonging to Advantech.



Prerounding Preparation and MDR With the Established EMR

Prerounding data gathering and MDRs have long been standardized with a structured script for reporting (Multimedia Appendix 3). Without *i*-Dashboard, NPs accessed the established EMR systems, including ICCA, HIS, LIS, and PACS, for data gathering through desktop computers.

MDRs took place outside the patient rooms. To facilitate situation awareness of other participants, NPs delivered the oral case presentation and data communication based on the structured script, including basic information, catheter placement and their duration, vital signs, laboratory data, medications, input or output and nutrition, critical values, major image findings, consultations, and other major events (Multimedia Appendix 3). The intensivist summarized active problems, solicited feedback from nurses, RTs, pharmacists, and dietitians, and, if needed, provided in-depth knowledge on the pathophysiology of the current patient condition. The goals of care were documented.

Prerounding Preparation and MDRs With i-Dashboard

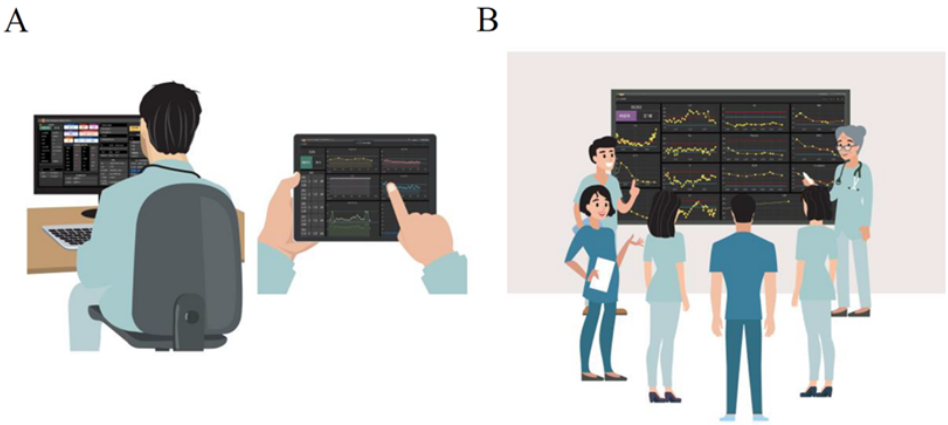
To ensure effective implementation of *i*-Dashboard, we prepared education materials (Multimedia Appendix 2) in the form of brief presentations for all the health care providers working in study units. Subsequently, the use of *i*-Dashboard was tested consistently for 4 weeks.

As shown in Figure 2, *i*-Dashboard serves as a platform to facilitate MDRs. NPs accessed *i*-Dashboard for data gathering through desktop computers. Instead of ~20 geographically fragmented windows and panels in our established EMR systems, at-a-glance presentations of highly relevant information were displayed on *i*-Dashboard.

During MDRs, all MDR participants gathered in front of a 55-inch 4K interactive touch screen, approximately 3 m away with clear sightlines. The touch screen allowed users to enter different pages of *i*-Dashboard using the finger to tap hyperlinks. The *i*-Dashboard served as a visualization aid for exchanging information and opinions. With the overview page in *i*-Dashboard, the NP carried out patient presentation and data communication. The NP accessed time-series data (eg, laboratory data) through the hyperlinks on the overview page. Time-series data can be rearranged on the basis of different time frames as requested. In addition, the RT, pharmacist, and dietitian could take turns operating *i*-Dashboard and use the RT-pharmacist-dietitian page to demonstrate valuable information of their professionals. Finally, the intensivist used *i*-Dashboard to facilitate bedside teaching. The goals were documented after consensus was reached.

XSL•FO
RenderX

Figure 2. i-Dashboard as a platform to facilitate multidisciplinary rounds. (A) Data access through i-Dashboard on different devices (eg, desktop computers and mobile platforms). (B) i-Dashboard displayed on wall-mounted large touch screens as a visualization aid during multidisciplinary rounds.



Data Collection

Two NPs who were not directly involved in MDRs and patient care were trained to audit the processes. The 2 NPs have 8 and 6 years of ICU experience, respectively. The 2 observers were temporarily exempted from clinical work during the study period. To ensure adequate training in the study methodology, personnel piloted data collection and evaluation of communication accuracy were performed by the 2 observers and supervised by the senior investigator (CHL) during a 4-week run-in period. The 2 observers performed in-field observation and audio recordings and audited the process together using a standardized form (Multimedia Appendix 4) to reduce the possibility of losing any useful information and to ensure the correct assessment. Before MDRs, observers measured the amount of time for prerounding data gathering by the NP using the built-in stopwatch app on mobile phones. Subsequently, observers arrived before the start of MDRs and refrained from participating in the discussion during MDRs. Clinical characteristics of patients, information on patient disease severity, and therapeutic interventions during MDRs and established care plans were collected through EMR.

Primary Outcomes

Primary outcomes of interest included time of data gathering before MDRs and communication accuracy during MDRs. The amount of time that NPs spent gathering clinical data before rounding was recorded. Communication accuracy was evaluated on the basis of the items listed on the standardized form (Multimedia Appendix 4) through direct field observation and audio recordings. Spoken information was compared with EMR data captured by screenshots taken prior to patient presentations. Data communication, including laboratory and nonlaboratory data, was considered inaccurate (ie, data misrepresentation) when the values or data were not correctly reported. For laboratory data communication, only abnormal laboratory data points (outside the reference ranges) were assessed. Laboratory misrepresentations were further classified into several categories as previously defined [8], including omission, old data, pending results, misinterpretation, and erroneous values.

Secondary Outcomes

Secondary outcomes included information exchange, using effective recommendations initiated by RTs, pharmacists, and dietitians as an index, and clinical satisfaction on *i*-Dashboard. The recommendations initiated by RTs, pharmacists, and dietitians, as exemplified in Multimedia Appendix 5, were considered effective on the condition that they were successfully adopted into the care plan documented in the EMR. Clinical satisfaction with *i*-Dashboard as an information management tool was investigated using Likert scale–based questionnaires (Multimedia Appendix 6) from previously validated survey instruments with minor modifications [12]. Questionnaire 1 was designed to capture the perceived efficiency, accuracy, and safety of the established EMR and *i*-Dashboard was implemented when used to prepare for data gathering and to assist MDRs. The responses were grouped into four dimensions: task productivity, task innovation, customer satisfaction, and management control. Questionnaire 2 was designed to identify intention to use and personal impact of *i*-Dashboard as a result of *i*-Dashboard implementation. At the immediate end of the study, the 2 surveys were administered in hard copy form to study participants. Each participant responded only once.

Sample Size Estimation and Statistical Analysis

Before the trial began, it was estimated that 12 MDRs would be observed per unit in a 2-week period. To estimate the required sample size, we used a cluster-randomized controlled trial design to account for the positive intraclass correlation expected among members of the same group or cluster. Observational pilot data for prerounding data gathering were collected. The results showed that the mean time difference between the use of the established EMR and *i*-Dashboard was 3 (variance 4) minutes. The intraclass correlation coefficient of this pilot study was 0.47. The corresponding estimations were treated as true parameters. Thus, enrolling approximately 144 patients during 24 unit-weeks would provide a power of 90% at a type I error rate of 0.05 to detect an intervention effect of 3 minutes between groups. Categorical variables were analyzed using the chi-square test or Fisher exact test as needed. Continuous variables and the Likert scale were analyzed using the Mann-Whitney *U* test. Statistical analyses were performed using R software (version 3.4.3; The R Foundation). A 2-tailed *P* value of <.05 was considered statistically significant.

XSL•FO
RenderX

Results

Participants

Between April 26 and July 18, 2021, there were 173 admissions to the 2 study units. A total of 90 multidisciplinary providers (Table 1) participated in MDRs for the 25 individual patients (Table 2); 78 MDRs were performed with the established EMR environment, whereas 91 MDRs were performed with *i*-Dashboard (Figure 3).

Table 1. Characteristics of the 90 multidisciplinary providers.

Variable	Value, n (%)
Sex	
Female	75 (83.3)
Male	15 (16.7)
Profession or discipline	
Physician	9 (10.0)
Nurse practitioner	6 (6.7)
Nurse	51 (56.7)
Respiratory therapist	20 (22.2)
Pharmacist	2 (2.2)
Dietitian	2 (2.2)
Intensive care unit experience (years)	
<1	3 (3.3)
1-2	20 (22.3)
3-4	15 (16.7)
5-9	31 (34.4)
>10	21 (23.3)

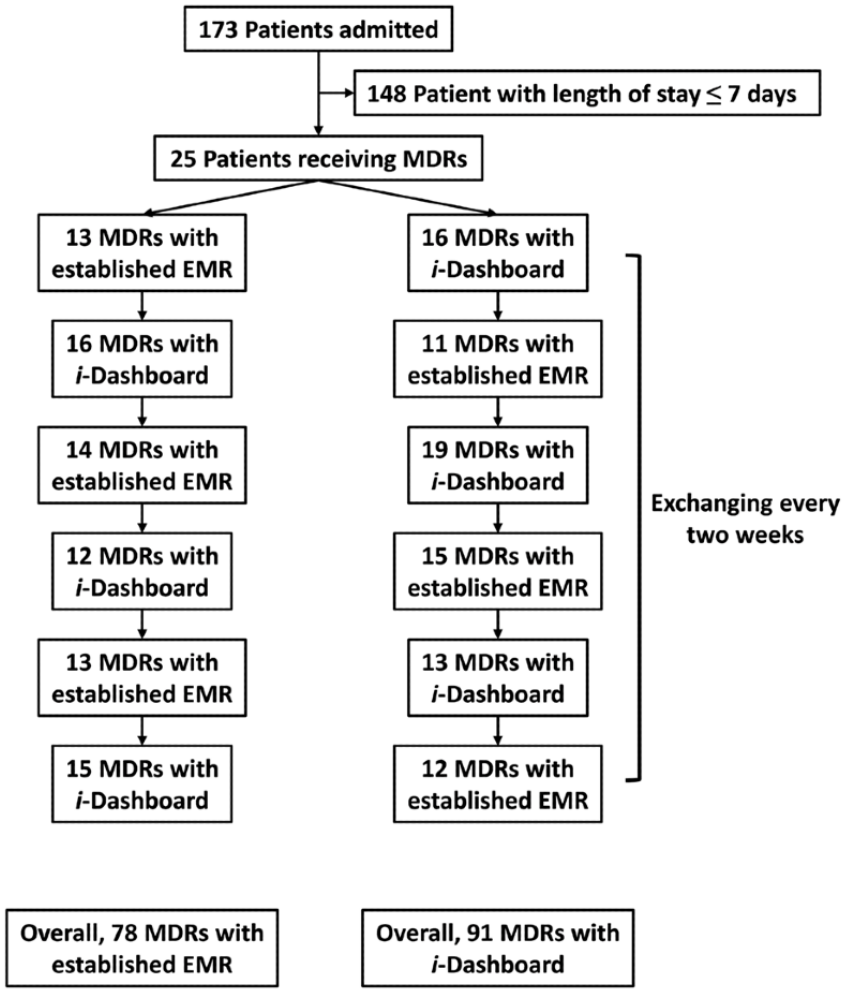
XSL•FO
RenderX

Table 2. Clinical characteristics of the 25 patients.

Variable	Value
Age (years), median (IQR)	70 (58-73)
Age distribution (years), n (%)	
<60	7 (28.0)
60-79	17 (68.0)
>80	1 (4.0)
Sex, n (%)	
Female	11 (44.0)
Male	14 (56.0)
Specialty, n (%)	
General surgery	8 (32.0)
Neurosurgery	11 (44.0)
Cardiovascular surgery	5 (2.0)
Trauma surgery	1 (4.0)
Type of admission, n (%)	
Medical	10 (40.0)
Scheduled surgical	7 (28.0)
Unscheduled surgical	8 (32.0)
Acute Physiology and Chronic Health Evaluation II score on admission, n (%)	
<15	1 (4.0)
15-34	21 (84.0)
>35	3 (12.0)
Mortality, n (%)	2 (8.0)

XSL•FO
RenderX

Figure 3. Study flowchart. EMR: electronic medical record, MDR: multidisciplinary round.



Primary Outcomes

Disease severity, in terms of severity scores, and therapeutic interventions at the day of MDRs were not different between the 2 groups (Table 3). The median time for prerounding data gathering was 10.4 (IQR 9.1-11.8) minutes and 4.6 (IQR 3.5-5.8) minutes per patient using the established EMR and *i*-Dashboard (P<.001; Table 3), respectively, indicating a reduction of 5.8 (95% CI 5.2-6.4) minutes when using *i*-Dashboard.

Regarding communication accuracy during MDRs (Table 3), data misrepresentations were significantly less frequent in MDRs with *i*-Dashboard (median 0, IQR 0-0) than with the established EMR environment (median 4, IQR 3-5; P<.001). In addition, both laboratory and nonlaboratory data misrepresentations were reduced using *i*-Dashboard compared with the established EMR. Among audited laboratory results, only one misrepresentation (0.2%) occurred among 577 data points with *i*-Dashboard. In contrast, 163 (32.3%) misrepresentations occurred in 505 data points with the established EMR environment (P<.001), and the majority (95.1%) of these misrepresentations were omissions (155 data points).

XSL•FO
RenderX

Table 3. Disease severity and therapeutic intervention at the moment of MDRs and outcomes with the established EMR environment and i-Dashboard.

Variable	Established electronic medical record (n=78)	*i*-Dashboard (n=91)	*P* value
Severity scoring, median (IQR)			
Modified Early Warning Score	6 (5-8)	7 (5-8)	.95
Simplified Acute Physiology Score II	52 (39-61)	50 (42-62)	.92
Sequential Organ Failure Assessment	8 (5-11)	7 (4-11)	.57
Therapeutic Intervention Scoring System-28	35 (32-38)	33 (30-39)	.19
Therapeutic intervention, n (%)			
Mechanical ventilation	72 (92.3)	86 (94.5)	.56
Vasoactive drug support	18 (23.1)	32 (35.2)	.08
Mechanical support	3 (3.8)	0 (0)	.10
Total parenteral nutrition	29 (37.2)	30 (33.0)	.57
Complicated wound management	24 (30.8)	23 (25.3)	.43
Dialysis-requiring renal failure	11 (14.1)	13 (14.3)	.73
Outcome			
Time spent on data gathering (minutes), median (IQR)	10.4 (9.1-11.8)	4.6 (3.5-5.8)	<.001
Data misrepresentation, median (IQR)	4 (3-5)	0 (0-0)	<.001
Laboratory data	2 (1-3)	0 (0-0)	<.001
Nonlaboratory data	2 (1-3)	0 (0-0)	<.001
Effective recommendations, n (%)			<.001
0	16 (20.5)	15 (16.5)	
1	41 (52.6)	34 (37.4)	
2	18 (23.1)	25 (27.5)	
3	3 (3.8)	17 (18.7)	

Secondary Outcomes

Regarding information exchange (Table 3), the proportions of 0 and 1 recommendations were lower in MDRs with *i*-Dashboard than with the established EMR, whereas the proportions of 2 and 3 recommendations were higher in MDRs with *i*-Dashboard than with the established EMR. The number of effective recommendations was significantly higher in MDRs with *i*-Dashboard than with the established EMR (P<.001).

A total of 76 health care providers (Table S1 in Multimedia Appendix 7) responded to the survey request of clinical satisfaction of *i*-Dashboard (response rate=84.4%). Grouping results of responses to Questionnaire 1 in term of task productivity, task innovation, customer satisfaction, and management control are shown in Table 4, and details are shown in Table S2 in Multimedia Appendix 7. Grouping results revealed that *i*-Dashboard was superior to the established EMR in task productivity (mean 15.91, SD 2.28 vs 14.14, SD 2.35; P<.001). Further, *i*-Dashboard was superior to established EMR in task innovation (12.11, SD 1.92 vs 10.41, SD 1.97; P<.001), customer satisfaction (16.68, SD 2.02 vs 15.62, SD 2.27; P=.002), and management control (16.75, SD 2.21 vs 15.03, SD 2.30; P<.001). These findings suggest that *i*-Dashboard outperformed the established EMR across the 4 dimensions.

Finally, survey responses to questionnaire 2 (Table S3 in Multimedia Appendix 7) suggested that these participants were willing to use *i*-Dashboard continuously in association with the enhancement of situation awareness, care plan development, and team participation, and with a reduction in workload and complexity.

XSL•FO
RenderX

Table 4. Grouping results of responses to questionnaire 1 in terms of task productivity, task innovation, customer satisfaction, and management control (n=76).

Question	Established electronic medical record, mean (SD)[a]	*i*-Dashboard, mean (SD)[a]	*P* value
Task productivity	14.14 (2.35)	15.91 (2.28)	<.001
Q1. _____ provides information catching up with condition changes.	3.68 (0.79)	4.17 (0.70)	
Q4. I get the information that I need in time using _____.	3.82 (0.69)	4.12 (0.59)	
Q5. I get the information that I need using _____ easily.	3.61 (0.80)	4.20 (0.69)	
Q9. _____ makes data gathering difficult.	3.04 (0.93)	3.42 (1.04)	
Task innovation	10.41 (1.97)	12.11 (1.92)	<.001
Q10. Data gathering with _____ was a mentally demanding task.	2.91 (1.00)	3.58 (1.07)	
Q14. Communication and opinion exchange in MDRs[b] is enhanced using _____.	3.75 (0.77)	4.25 (0.66)	
Q15. Developing care plans relies on joint decisions by team members using _____.	3.75 (0.73)	4.28 (0.65)	
Customer satisfaction	15.62 (2.27)	16.68 (2.02)	.002
Q2. _____ provides information that meets my demand for following MDRs.	4.12 (0.63)	4.29 (0.61)	
Q3. _____ provides me sufficient information for patient care.	4.00 (0.71)	4.16 (0.63)	
Q6. I am satisfied with the accuracy of the data using _____.	3.87 (0.68)	4.03 (0.63)	
Q13. _____ makes me fully understanding the situation and goal of each patient.	3.63 (0.83)	4.21 (0.60)	
Management control	15.03 (2.30)	16.75 (2.21)	<.001
Q7. The information presented by _____ is clear.	3.78 (0.70)	4.25 (0.61)	
Q8. The information presented in the format of _____ is effective and useful.	3.67 (0.76)	4.29 (0.63)	
Q11. The information presented using _____ during MDRs was accurate.	3.87 (0.62)	4.07 (0.62)	
Q12. The presentation of patient information during MDRs using _____ was organized.	3.71 (0.71)	4.15 (0.60)	

[a]Values in Q9 and Q10 were calculated by reverse scoring.

[b]MDR: multidisciplinary round.

Discussion

Principal Findings

The *i*-Dashboard was developed as a structured, process-oriented information platform for MDRs, where the efficiency of data retrieval, fidelity of data communication, and satisfaction of interdisciplinary communication are all requisites. Through a comprehensive evaluation of the performance of *i*-Dashboard, we found that under similar disease complexity, *i*-Dashboard may increase efficiency in prerounding data gathering compared to the established EMR. More importantly, displaying *i*-Dashboard on large touch screens in MDRs may enhance communication accuracy, information exchange, and clinical satisfaction.

Clinical Aspects

Information overload has been a severe problem in the ICU [18]. Critical care providers express frustration with the difficulty in organizing data, especially quantitative dynamic data (eg, deteriorating serum creatinine levels during acute kidney injury), and become overwhelmed by data overload [9,12,13,19]. In time-sensitive care environments, such as the emergency department and the ICU, visualization may provide information that can be readily perceived, easily recognized, and processed expeditiously into inferences [20]. In addition, visualization through dashboards may provide memory aids [11,13,20]. Implementing visualization dashboards in the clinical setting may improve data display and reduce cognitive overload among clinicians [7,9,11,13,21].

This study proposes the notion that large-screen visualization dashboards may improve data communication and information exchange during ICU MDRs. In the emergency department, large-screen visualization dashboards help health care providers find the desired information without wasting time [20,21]. In our study, *i*-Dashboard displayed on large, wall-mounted monitors could present well-organized data visually for the multidisciplinary care team during MDRs and thus avoid data misrepresentations through verbal communication. Compared with small computer monitors, a large screen display effectively prevents unequal access to data and seems more likely to

XSL•FO
RenderX

establish consensus. Thus, effective visualization through *i*-Dashboard displayed on large screens contributes to a better perception of information for decision-making by the multidisciplinary care team. When considering the user experience, dashboard visualization may improve the perception and comprehension of patient-level information [20], thereby removing barriers that hinder information exchange. Additionally, the concern regarding the prohibitive cost for large interactive touch screens [14] has been greatly attenuated.

Comparison With Previous Work

Visualization dashboards can inform decision-making and support behavior change in public health and health care services [22-25]. A recent review of the literature suggests that the strength of evidence on the effect of ICU visualization dashboards remains low [9]. Of the 4 available randomized controlled trials, only Pickering et al [12] found a significant improvement compared to the pre-existing EMR environment. Time spent for prerounding data gathering efforts is practical to evaluate information tools on MDRs [7]. In their study, participants who had access to the AWARE dashboard significantly decreased the data gathering time from 12 to 9 minutes per patient (P<.03). In this study, we estimated a reduction of ~60 mouse clicks per patient using *i*-Dashboard versus the established EMR before the trial. For experienced NPs, this improvement can be translated to a 6-minute reduction, compatible with what has been observed (5.8 minutes) during the trial. Health care providers can reduce cognitive fatigue in the data extraction process and pay attention to more productive information.

The ICU is a dynamic environment in which multiple information pathways and personnel interactions facilitate patient care. Specially trained health care professionals in the ICU rely on interdisciplinary communication to make effective clinical decisions. Reliance on a single individual in patient data communication during MDRs represents a coping strategy for an EMR system that does not automatically provide an effective visualization display of the data needed [8]. Consistent with previous studies [8], we found that verbally shared communication of patient data during MDRs was prone to errors and inaccuracies, and most of the laboratory misrepresentations were omissions. Regardless of laboratory or nonlaboratory data, these misrepresentations could be almost eliminated using *i*-Dashboard as a visualization aid for MDRs.

While structured presentation using checklists and explicit definitions of each health care provider's role may facilitate MDRs, allied health care provider perceptions of not being valued by rounding intensivists have been recognized as an unfavorable factor that impairs the productivity of MDRs. In MDRs with *i*-Dashboard, participants of different professions had a space and time interval of their own, thereby enhancing the sense of participation and collaboration. *i*-Dashboard may help the RT, pharmacist, and dietitian focus on expressing their thoughts explicitly, as other participants could obtain numerical information from the visualization aid. Their inputs were thus more likely to be valued and adopted by the multidisciplinary care team. The positive impacts of *i*-Dashboard on objective outcomes were corroborated by the increased perception of situation awareness and team participation in participants, as revealed in the questionnaire results.

We found that after the study period, health care providers in study units started to propose novel ideas that might improve the usability of *i*-Dashboard (eg, rearranging the layout and increasing laboratory items). A recent study has demonstrated valuable experience regarding the evolution process of an emergency department dashboard, which has undergone several significant revisions to respond to feedback from users [21]. Currently, we are developing the second version of *i*-Dashboard.

Limitations

Our findings must be interpreted within the context of the study limitations. First, traditional study outcomes in critical care, such as mortality and length of stay, were not evaluated in this study. No conclusion can be achieved regarding the effect of *i*-Dashboard on patient outcomes. Further studies are warranted in this respect. Second, the designers of *i*-Dashboard were part of the team that conducted the study and assessed the outcomes. Successful implementation of dashboards greatly relies on user experience. However, the participation of this paper's authors in the development process potentially leads to biases in assessing outcomes [26]. Finally, *i*-Dashboard was custom-developed with reference to our established EMR environment. Therefore, it is difficult to extrapolate our study findings directly into ICUs of other hospitals, possibly limiting their generalizability. Nevertheless, visualization dashboards are intended to reduce the time spent on the data gathering process and improve situation awareness and navigation [11,21]. The promising results of *i*-Dashboard obtained in a mature MDR environment suggest that these design concepts may help develop or modify visualization dashboards in the ICU more applicably for MDRs through technological advancements.

Conclusions

In conclusion, we developed *i*-Dashboard as an information management platform for MDRs. The implementation of *i*-Dashboard can increase efficiency in prerounding data gathering. As a visualization aid, *i*-Dashboard displayed on large screens enhances communication accuracy and information exchange during MDRs. Establishing care team–designed visualization dashboards as an integrated information platform may reinforce the communication quality of MDRs, thus potentially improving the workflow process in the ICU.

Acknowledgments

We appreciate Prof Yu Shyr, Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, for his help in study design, Mr Cheng-Ping Lai, Department of Information Technology, for his excellent technical support, Ms Yu Miao, Department of Statistics, for her contribution in data analysis, and Ms Yu-Fang Hsieh, Ms Pei-Jung Wu, Ms Li-Chin Hsiao,

XSL•FO
RenderX

Ms Tzu-Yun Chiu, and Ms Kuan-Ling Chin, Department of Nursing, for their laborious contribution to this work. This work is supported by grants from the Ministry of Science and Technology, Executive Yuan, Taiwan (MOST 109-2634-F-006-023 to M-RS).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist.
[PDF File (Adobe PDF File), 1391 KB - jmir_v24i5e35981_app1.pdf]

Multimedia Appendix 2
Architecture of i-Dashboard.
[PDF File (Adobe PDF File), 2433 KB - jmir_v24i5e35981_app2.pdf]

Multimedia Appendix 3
Structured script for data gathering and patient presentation.
[PDF File (Adobe PDF File), 216 KB - jmir_v24i5e35981_app3.pdf]

Multimedia Appendix 4
Standardized form for evaluating communication accuracy.
[PDF File (Adobe PDF File), 241 KB - jmir_v24i5e35981_app4.pdf]

Multimedia Appendix 5
Recommendations initiated by respiratory therapists, pharmacists and dietitians.
[PDF File (Adobe PDF File), 535 KB - jmir_v24i5e35981_app5.pdf]

Multimedia Appendix 6
Questionnaires.
[PDF File (Adobe PDF File), 416 KB - jmir_v24i5e35981_app6.pdf]

Multimedia Appendix 7
Questionnaire results.
[PDF File (Adobe PDF File), 332 KB - jmir_v24i5e35981_app7.pdf]

References

1. Halpern NA, Pastores SM. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. Crit Care Med 2010 Jan;38(1):65-71. [doi: 10.1097/CCM.0b013e3181b090d0] [Medline: 19730257]
2. Vincent J, Marshall JC, Namendys-Silva SA, François B, Martin-Loeches I, Lipman J, ICON investigators. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. Lancet Respir Med 2014 May;2(5):380-386. [doi: 10.1016/S2213-2600(14)70061-X] [Medline: 24740011]
3. Yoo EJ, Edwards JD, Dean ML, Dudley RA. Multidisciplinary critical care and intensivist staffing: results of a statewide survey and association with mortality. J Intensive Care Med 2016 Jun;31(5):325-332. [doi: 10.1177/0885066614534605] [Medline: 24825859]
4. Kim MM, Barnato AE, Angus DC, Fleisher LA, Fleisher LF, Kahn JM. The effect of multidisciplinary care teams on intensive care unit mortality. Arch Intern Med 2010 Feb 22;170(4):369-376 [FREE Full text] [doi: 10.1001/archinternmed.2009.521] [Medline: 20177041]
5. Lane D, Ferri M, Lemaire J, McLaughlin K, Stelfox HT. A systematic review of evidence-informed practices for patient care rounds in the ICU. Crit Care Med 2013 Aug;41(8):2015-2029. [doi: 10.1097/CCM.0b013e31828a435f] [Medline: 23666096]
6. Leonard MW, Frankel AS. Role of effective teamwork and communication in delivering safe, high-quality care. Mt Sinai J Med 2011;78(6):820-826. [doi: 10.1002/msj.20295] [Medline: 22069205]
7. Gurses AP, Xiao Y. A systematic review of the literature on multidisciplinary rounds to design information technology. J Am Med Inform Assoc 2006;13(3):267-276 [FREE Full text] [doi: 10.1197/jamia.M1992] [Medline: 16501176]

XSL•FO
RenderX

8. Artis KA, Dyer E, Mohan V, Gold JA. Accuracy of laboratory data communication on ICU daily rounds using an electronic health record. Crit Care Med 2017 Feb;45(2):179-186 [FREE Full text] [doi: 10.1097/CCM.0000000000002060] [Medline: 27655323]
9. Waller RG, Wright MC, Segall N, Nesbitt P, Reese T, Borbolla D, et al. Novel displays of patient information in critical care settings: a systematic review. J Am Med Inform Assoc 2019 May 01;26(5):479-489 [FREE Full text] [doi: 10.1093/jamia/ocy193] [Medline: 30865769]
10. De Georgia MA, Kaffashi F, Jacono FJ, Loparo KA. Information technology in critical care: review of monitoring and data acquisition systems for patient care and research. ScientificWorldJournal 2015;2015:727694 [FREE Full text] [doi: 10.1155/2015/727694] [Medline: 25734185]
11. Khairat SS, Dukkipati A, Lauria HA, Bice T, Travers D, Carson SS. The impact of visualization dashboards on quality of care and clinician satisfaction: integrative literature review. JMIR Hum Factors 2018 May 31;5(2):e22 [FREE Full text] [doi: 10.2196/humanfactors.9328] [Medline: 29853440]
12. Pickering BW, Dong Y, Ahmed A, Giri J, Kilickaya O, Gupta A, et al. The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: a pilot step-wedge cluster randomized trial. Int J Med Inform 2015 May;84(5):299-307. [doi: 10.1016/j.ijmedinf.2015.01.017] [Medline: 25683227]
13. Wright MC, Borbolla D, Waller RG, Del Fiol G, Reese T, Nesbitt P, et al. Critical care information display approaches and design frameworks: a systematic review and meta-analysis. J Biomed Inform 2019;100S:100041 [FREE Full text] [doi: 10.1016/j.yjbinx.2019.100041] [Medline: 34384575]
14. Morrison C, Jones M, Blackwell A, Vuylsteke A. Electronic patient record use during ward rounds: a qualitative study of interaction between medical staff. Crit Care 2008;12(6):R148 [FREE Full text] [doi: 10.1186/cc7134] [Medline: 19025662]
15. Aslakson RA, Wyskiel R, Shaeffer D, Zyra M, Ahuja N, Nelson JE, et al. Surgical intensive care unit clinician estimates of the adequacy of communication regarding patient prognosis. Crit Care 2010;14(6):R218 [FREE Full text] [doi: 10.1186/cc9346] [Medline: 21114837]
16. 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] [Medline: 22209829]
17. Roumeliotis N, Parisien G, Charette S, Arpin E, Brunet F, Jouvet P. Reorganizing care with the implementation of electronic medical records: a time-motion study in the PICU. Pediatr Crit Care Med 2018 Apr;19(4):e172-e179. [doi: 10.1097/PCC.0000000000001450] [Medline: 29329162]
18. Manor-Shulman O, Beyene J, Frndova H, Parshuram CS. Quantifying the volume of documented clinical information in critical illness. J Crit Care 2008 Jun;23(2):245-250. [doi: 10.1016/j.jcrc.2007.06.003] [Medline: 18538218]
19. Wright M, Dunbar S, Macpherson B, Moretti E, Del Fiol G, Bolte J, et al. Toward designing information display to support critical care. A qualitative contextual evaluation and visioning effort. Appl Clin Inform 2016 Oct 05;7(4):912-929 [FREE Full text] [doi: 10.4338/ACI-2016-03-RA-0033] [Medline: 27704138]
20. Franklin A, Gantela S, Shifarraw S, Johnson TR, Robinson DJ, King BR, et al. Dashboard visualizations: supporting real-time throughput decision-making. J Biomed Inform 2017 Jul;71:211-221 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.024] [Medline: 28579532]
21. Yoo J, Jung KY, Kim T, Lee T, Hwang SY, Yoon H, et al. A real-time autonomous dashboard for the emergency department: 5-year case study. JMIR Mhealth Uhealth 2018 Nov 22;6(11):e10666 [FREE Full text] [doi: 10.2196/10666] [Medline: 30467100]
22. Ivanković D, Barbazza E, Bos V, Brito Fernandes Ó, Jamieson Gilmore K, Jansen T, et al. Features constituting actionable COVID-19 dashboards: descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. J Med Internet Res 2021 Feb 24;23(2):e25682 [FREE Full text] [doi: 10.2196/25682] [Medline: 33577467]
23. Gremyr A, Andersson Gäre B, Greenhalgh T, Malm U, Thor J, Andersson A. Using complexity assessment to inform the development and deployment of a digital dashboard for schizophrenia care: case study. J Med Internet Res 2020 Apr 23;22(4):e15521 [FREE Full text] [doi: 10.2196/15521] [Medline: 32324143]
24. Bos VLLC, Jansen T, Klazinga NS, Kringos DS. Development and actionability of the Dutch COVID-19 dashboard: descriptive assessment and expert appraisal study. JMIR Public Health Surveill 2021 Oct 12;7(10):e31161 [FREE Full text] [doi: 10.2196/31161] [Medline: 34543229]
25. Zheng S, Edwards JR, Dudeck MA, Patel PR, Wattenmaker L, Mirza M, et al. Building an interactive geospatial visualization application for national health care-associated infection surveillance: development study. JMIR Public Health Surveill 2021 Jul 30;7(7):e23528 [FREE Full text] [doi: 10.2196/23528] [Medline: 34328436]
26. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 09;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]

Abbreviations

EMR: electronic medical record

XSL•FO
RenderX

HIS: Hospital Information Systems
ICCA: IntelliSpace Critical Care & Anesthesia information system
ICU: intensive care unit
LIS: Laboratory Information System
MDR: multidisciplinary round
NP: nurse practitioner
PACS: Picture Archiving and Communication System
RT: respiratory therapist

Edited by T Leung; submitted 02.01.22; peer-reviewed by S Khairat, M Wright, H He; comments to author 30.01.22; revised version received 20.02.22; accepted 25.04.22; published 13.05.22.

Please cite as:
Lai CH, Li KW, Hu FW, Su PF, Hsu IL, Huang M, Huang Y, Liu PY, Shen MR
Integration of an Intensive Care Unit Visualization Dashboard (i-Dashboard) as a Platform to Facilitate Multidisciplinary Rounds: Cluster-Randomized Controlled Trial
J Med Internet Res 2022;24(5):e35981
URL: https://www.jmir.org/2022/5/e35981
doi:10.2196/35981
PMID:35560107

©Chao-Han Lai, Kai-Wen Li, Fang-Wen Hu, Pei-Fang Su, I-Lin Hsu, Min-Hsin Huang, Yen-Ta Huang, Ping-Yen Liu, Meng-Ru Shen. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.05.2022. 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.

XSL•FO
RenderX

Original Paper

Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework

Bettina M Zimmermann[1,2*], DPhil; Theresa Willem[1,3*], MA; Carl Justus Bredthauer[1], BA; Alena Buyx[1], DrMed

[1]Institute of History and Ethics in Medicine, School of Medicine, Technical University of Munich, Munich, Germany

[2]Institute for Biomedical Ethics, University of Basel, Basel, Switzerland

[3]Department of Science, Technology and Society, School of Social Sciences and Technology, Technical University of Munich, Munich, Germany

[*]these authors contributed equally

Corresponding Author:
Theresa Willem, MA
Institute of History and Ethics in Medicine, School of Medicine
Technical University of Munich
Prinzregentenstraße 68
Munich, 81675
Germany
Phone: 49 17662315637
Email: theresa.willem@tum.de

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/9/e40848

Abstract

Background: Social media recruitment for clinical studies holds the promise of being a cost-effective way of attracting traditionally marginalized populations and promoting patient engagement with researchers and a particular study. However, using social media for recruiting clinical study participants also poses a range of ethical issues.

Objective: This study aims to provide a comprehensive overview of the ethical benefits and risks to be considered for social media recruitment in clinical studies and develop practical recommendations on how to implement these considerations.

Methods: On the basis of established principles of clinical ethics and research ethics, we reviewed the conceptual and empirical literature for ethical benefits and challenges related to social media recruitment. From these, we derived a conceptual framework to evaluate the eligibility of social media use for recruitment for a specific clinical study.

Results: We identified three eligibility criteria for social media recruitment for clinical studies: information and consent, risks for target groups, and recruitment effectiveness. These criteria can be used to evaluate the implementation of a social media recruitment strategy at its planning stage. We have discussed the practical implications of these criteria for researchers.

Conclusions: The ethical challenges related to social media recruitment are context sensitive. Therefore, social media recruitment should be planned rigorously, taking into account the target group, the appropriateness of social media as a recruitment channel, and the resources available to execute the strategy.

(J Med Internet Res 2022;24(5):e31231) doi:10.2196/31231

KEYWORDS

social media; clinical studies; clinical trials; ethics; recruitment

Introduction

Effective patient recruitment has been one of the most cited barriers to clinical studies [1,2]. Up to 60% of trials are delayed or canceled because of a lack of enrollment [3-5]. In the recent past, social media recruitment has been successfully used for different clinical studies (eg, smoking cessation [6], type 1 diabetes [7], and HIV [8]), which has raised hopes for the improvement of the research process, quality, and efficiency. The COVID-19 pandemic has further enhanced the use of social media platforms.

XSL•FO
RenderX

However, researchers are calling for a more thorough evaluation of recruitment effectiveness, quality, and cost-effectiveness [9], as well as their ethical, legal, and social implications. Recent publications have raised doubts about whether social media can be used for recruitment purposes while preserving fundamental cornerstones of research ethics and biomedical ethics, as reflected in the US Health Insurance Portability and Accountability Act, the General Data Protection Regulation of the European Union, the Declaration of Helsinki, and the Belmont Report, to name a few examples [3,10-15].

Few scholars have examined the unique ethical, legal, and social issues arising from social media recruitment at a metalevel. Gelinas et al [3] have provided an overview of the ethical issues related to social media recruitment in research. They mainly addressed concerns related to user privacy and investigator transparency and provided practical recommendations for internal review boards and investigators. In particular, they claimed that social media recruitment should follow the same standards as traditional recruitment strategies. In this review, we aim to continue this work but take a broader view of the ethical issues that might arise with regard to social media recruitment.

We explore the uniqueness of social media recruitment compared with more traditional forms of patient recruitment. In doing so, we consider the novel architectures of social media platforms and the possibilities they offer for recruitment communication compared with prior web-based and offline one-to-one and one-to-many communication channels, such as email recruitment or billboard recruitment. Our analysis is based on the benefits and challenges of complex community-based communication opportunities that unfold based on different privacy settings. We also examine the application of machine learning techniques for predictive analytic purposes of user behavior based on central databases [16].

We aim to target researchers involved in clinical studies who are considering incorporating social media into their recruitment strategies. Specifically, we seek to provide a comprehensive overview of the ethical issues to be considered from conceptual and empirical perspectives (including the potential benefits and risks) and provide practical recommendations on how to take these issues into account when using social media as a recruitment tool. Therefore, our recommendations are formulated with a view toward practical application.

Methods

We conducted a structured, nonsystematic review of the empirical evidence available for social media recruitment and its links to ethical challenges. First, we broadly reviewed the normative and empirical scientific literature on social media recruitment. Then, we identified relevant papers based on their abstracts through searches in interdisciplinary databases (PubMed [MEDLINE], Web of Science, Google Scholar, and the university library catalogs of the Technical University Munich and Ludwig-Maximilian University Munich) using the initial keywords *social media* AND *recruitment* AND *clinical trial**. We also screened the references of the relevant articles. From the publications relevant to social media recruitment, we collected all relevant ethical issues and grouped them thematically according to the principles of biomedical ethics [17], namely, autonomy, justice, nonmaleficence, and beneficence. We then checked the ethical issues for conceptual completeness using 2 well-established normative frameworks for research ethics as conceptual guidance [18,19]. These frameworks guide important concepts of state-of-the-art biomedical and ethical research, such as privacy, informed consent, specific protection of the vulnerable, and other potential risks relevant in this context. After identifying conceptually relevant ethical benefits and challenges, we performed additional literature searches to gain more focused insights into the available empirical literature on these benefits and challenges related to social media recruitment for clinical studies. The details of these searches are available in Multimedia Appendix 1.

We present our results in 3 thematic chapters that investigate the benefits of social media recruitment for clinical studies and the associated ethical challenges (part A). As ethical benefits and challenges are highly context sensitive and require a thorough risk-benefit analysis for each clinical study, we used our findings to identify 3 dimensions that should be considered in an ethical assessment of any clinical study considering social media recruitment (part B). In the *Discussion* section, we analyze the practical implications of these dimensions in the context of clinical studies.

Results

Part A: Benefits and Challenges of Social Media Recruitment in Clinical Studies

Trust, Transparency, and Autonomy

Benefit: Promoting Trust, Transparency, and Autonomy

Conceptual and empirical research has encouraged the claim that social media can promote trust, transparency, and autonomy in research studies. For instance, the options for bilateral and multilateral interactions on social media enable participants to learn about the results of the clinical study in which they participate [20,21], which increases trust and transparency [22,23]. This allows for an individually adapted level of engagement between participants and researchers “in an era where the patients are collaborators and there is a continuum of need from paternalism to complete autonomy” [24]. Thus, when intended as an instrument to improve autonomy, social media provides an opportunity to promote patient empowerment [21]. In the realm of patient-led research and citizen science, social media can serve as a platform to bring researchers, patients, and other stakeholders together and foster collaboration [25,26]. This includes researchers being transparent about what data are collected and asking for feedback on the study results, thereby encouraging patient engagement [27]. As health data become increasingly accessible to individual patients outside the clinical setting, this is particularly important. However, data-rich medicine also gives rise to challenges for health care professionals and patients alike, such as supporting digital data practices or contextualizing them meaningfully [28]. In-depth and continuous exchanges between participants and researchers

XSL•FO
RenderX

are needed to promote trust, transparency, and autonomy, which could be included in a study's recruitment process.

Recruitment process reports suggest that using social media improves autonomous decision-making with regard to study participation by offering the possibility of multiple contact points over time, which reduces time pressure and supports informed decision-making [23,29]. Moreover, information can be presented multimodally (visual, aural, or tactile), which could improve the understanding of study-related information [30]. Reports of users sharing study-related content on their social media accounts, as well as analyses of responses and reactions on social media, have empirically demonstrated how perceived trust and transparency can be fostered on social media [31,32]. However, there are several ethical challenges, including issues related to the tension between information, nudging, and persuasion (refer to the *Challenge: Information, Persuasion, and Nudging* section), informed consent issues (refer to the *Challenge: Informed Consent* section), privacy and data security issues (refer to the *Challenge: Privacy and Data Security* section), and low digital literacy (refer to the *Challenge: Digital Literacy* section).

Challenge: Information, Persuasion, and Nudging

The boundaries between information and persuasion, as well as between education and advertisements, are ambiguous. Information processing has a direct effect on cognitive and emotional responses, motivation, and persuasion to action [33]. The concept of nudging people into beneficial behaviors [34] provides a potential framework that ethically justifies a certain level of persuasion, as long as it serves the best interest of the person. The nudging concept has been discussed extensively—and controversially—in the context of public health [35-38], informed consent in the clinic [39,40], and when providing incentives for research participation. More recently, it has also been discussed in terms of promoting privacy-preserving behavior in social media users [41-43]. Although the definition and scope of nudges are subject to debate, we refer to nudging as used by VanEpps et al [44], who suggested 3 forms of interventions to nudge research participation. The first form is simply *providing information* about the study, which is disadvantageous as it might lead to a lower recruitment accrual than other forms of nudging. Moreover, information can be provided in more or less persuasive ways, and what is perceived as appropriate information provision is highly context dependent. In any event, transparency regarding social media–related activities and interventions is an important prerequisite for any social media activity related to a clinical study [15] (see also *Challenge: Informed Consent* section).

The second form of nudging is *choice architecture.* For example, default choices might increase recruitment effectiveness but lead to informed consent issues that are unacceptable in the context of clinical studies. However, a recent study among surrogate decision-makers in an intensive care unit setting found no statistical difference between the 2 offers of choice architecture or any evidence of undue and unjust inducements [45].

The third form of nudging includes *monetary or other incentives.* This might lead to undue or unjust inducement and needs to be assessed in a context-specific manner [46,47]. Social nudging in the form of rewarding goal attainment has shown positive effects on individuals' willingness to act in a group's interest in the context of vaccine uptake [48]. However, empirical evidence concerning the effectiveness of these forms of nudging in the context of clinical study recruitment is limited and requires further investigation.

There has been little discussion concerning the implications of nudging and ethical assessments between informing users of the existence of a clinical study and persuading them to participate through monetary incentives or persuasive language. This needs further evaluation, as social media recruitment for clinical studies potentially combines impersonal communication methods with individual decision-making for medical interventions.

Challenge: Informed Consent

In the context of social media recruitment, informed consent processes must consider (1) when consent is required, (2) in what instances digital consent is sufficient, and (3) what educational and administrative hurdles are necessary to ensure that consent is informed [49].

Regarding the *requirement of consent*, Gelinas et al [3] called for specific consent when a participant's social network was used for further recruitment, as such methods might reveal private health-related information to a participant's social network. This consent acknowledges the context-specific adjustment of the informed consent process in different types of social media recruitment activities.

Regarding *digital consent*, empirical studies have suggested that web-based consent is potentially problematic as users are already used to agreeing to terms and conditions without informing themselves about the details [50]. Moreover, the informed consent processes that are applied vary depending on the study design. In web-based survey studies, consent is obtained directly on the web after the user is redirected from the social media platform to an external survey or intervention website [51,52]. In contrast, in randomized controlled trials, consent is typically sought offline after checking the eligibility criteria and before study enrollment [53,54]. Thus, current informed consent practices in medical studies using social media recruitment focus on consent *after* initial recruitment on social media. However, several authors have lamented the lack of guidelines concerning informed consent processes during or before recruitment for clinical trials via social media [3,55].

Regarding *educational and administrative hurdles*, the effectiveness of electronic forms of consent has been compared empirically with traditional paper-based forms. No universal best practices for e-consent have emerged [56,57]. Moreover, although previous studies have compared different forms of consent [58,59], there is a lack of studies that empirically examine the consent process in the context of social media recruitment for clinical studies.

XSL•FO
RenderX

Challenge: Privacy and Data Security

An extensive body of literature has investigated the complex and multifaceted concept of privacy; however, none of the existing definitions are universally applicable [60,61]. Here, we refer to the concept of *information privacy*, which is understood as a subset of privacy in general (refer to the studies by Smith et al [60] and Bélanger and Crossler [62] for comprehensive reviews on information privacy). This paper follows the definition of privacy proposed by Bélanger and Crossler [62]: the "desire of individuals to control or have some influence over data about themselves." Social media recruitment might conflict with this desire, for instance, when researchers or other potential participants tag individuals on social media recruitment posts without their permission and thereby unwillingly link them to a specific study or disease on a public platform [3,63]. Most importantly, recruiting researchers and other social media users may unknowingly cause such privacy violations (see the *Challenge: Digital Literacy* section). As with any other form of data collection, the researcher is responsible for avoiding the disclosure or loss of information collected on social media in connection with a clinical study [15]. However, researchers do not have control over the data shared on these platforms, making data management potentially challenging [49].

Many social media platforms offer features that advertise to a defined target population and use inaccessible algorithms to select them. For instance, Facebook's proprietary algorithm uses machine learning to infer the advertisements that should be displayed based on users' previous behavior [64,65]. A recent study showed that feeding machine learning algorithms with user data risks unforeseeable correlations that might be misused for predictive analytics [66]. Mühlhoff [67] argued that these algorithms have the potential to not only disclose details about a user's future behavior but also estimate details that users have not disclosed about themselves based on combinations of known data points (eg, stated preferences, demographics, and relations with other users). Thus, a social media platform can learn potentially sensitive information that a user does not want to disclose, posing a privacy challenge. Regulations for these procedures are lacking [68], as is empirical evidence of the ethical and social implications of these potentially problematic privacy issues.

Challenge: Digital Literacy

Another challenge affecting social media's potential to improve autonomy is the lack of digital literacy among social media users, defined as the skills and resources that users need to successfully navigate digital environments [69-71]. Individuals with low digital literacy might be at a greater risk of stigmatization or involuntary violation of their own or others' privacy. Furthermore, low digital literacy, often in combination with a lack of engagement on social media, may prevent individuals from finding and participating in clinical studies that mostly or solely use social media recruitment [31,71,72]. This may cause ethical challenges regarding the equality of access to clinical studies (see the *Challenge: Equality of Access* section). However, the importance of this issue depends on the population being targeted, as a younger population might be best reached via social media.

Empirical evidence suggests that many social media users lack the skills to self-assess potential risks and harms connected to their social media activities, especially when related to their health status [73-75]. Surveys with health and information technology professionals have found that low digital literacy is commonly perceived as causing ethical challenges when social media is used in the context of participatory health applications. Important concerns include accidental sharing of sensitive data and a limited understanding of what the data on social media are used for [76] (see the *Challenge: Information, Persuasion, and Nudging* section).

Justice and Nonmaleficence

Benefit: Including Marginalized Groups by Accessing Hard-to-Reach Populations

Ethnic minorities and other marginalized groups are often underrepresented in clinical studies, which causes the benefits of medical research to be unequally distributed [77]. This raises important ethical issues concerning discrimination and equitable access to care. First, there are significant differences in the reactions to medical procedures between subgroups of a population, and the underrepresentation of certain groups in clinical studies may lead to negative consequences for the safety and efficacy evidence in underrepresented ethnic groups [78]. Second, participation in clinical studies might also lead to better medical outcomes for study participants [79]. Third, the unwarranted exclusion of minorities from clinical studies represents a form of epistemic injustice—excluding certain populations from generating knowledge may lead to a biased, ungeneralizable body of knowledge [80].

A reason for the unequal representation in clinical studies is the difficulty in reaching certain groups. Researchers are usually aware of problems related to unequal representation and often cite issues when recruiting certain subgroups within a target group [81]. These hard-to-reach populations are often from socially disadvantaged groups, such as those with low socioeconomic status, ethnic minorities, or older adults [2,52,82].

Social media may help tackle these issues by increasing the ability of researchers to recruit hard-to-reach target populations. Owing to the potential of social media to alleviate the problems of unequal representation in clinical studies, Caplan and Friesen [81] even discussed a "duty to tweet." They stated that researchers should reduce inequalities in clinical studies and posited that targeted social media recruitment might be a valuable tool to meet this obligation.

Empirical assessments suggest that social media recruitment is effective in recruiting certain populations that are difficult to reach using traditional methods. Specifically, studies have shown that social media recruitment can overcome linguistic and educational barriers and reach immigrants and low-education and low-income populations [83,84]. In addition, social media has been useful in recruiting low-prevalence populations [85-87] and young individuals [15,88]. Moreover, social media recruitment has been successfully used to target gay couples for HIV and hepatitis B and C interventions [32,89]. As these

XSL•FO
RenderX

considerations are target group–specific, they must be assessed separately for each study.

Social media offers researchers and participants a low-threshold opportunity to engage with research. This is relevant from the perspective of marginalized groups as it provides value and allows for low-threshold access to potentially beneficial care [79].

Challenge: Stigmatization of Vulnerable Populations

Although vulnerability originally referred to the limited ability of persons to give informed consent (Department of Health, Education, and Welfare and National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research), the term was recently extended to include a more general definition and was applied to all those "incapable of protecting their own interest" [19]. Although the first definition tends to define vulnerability too narrowly, the second tends to define vulnerability too broadly, such that all human beings could be assigned to one vulnerable group or another [90]. To avoid this, vulnerability (meaning the potential to cause disproportional wrong to potential research participants [91]) should be assessed on a case-by-case basis when planning a research study. Furthermore, uncovered vulnerabilities should be re-examined situationally, as not every individual with a certain characteristic is vulnerable to the same extent under the same circumstances. This includes planning ahead for a patient's potential situations of vulnerability [92-95]. Clinical studies targeting patients with characteristics that render them vulnerable must consider thorough informed consent processes and options to withdraw consent (refer to the *Challenge: Informed Consent* section). Consequently, both the vulnerability status and informed consent process are highly dependent on the respective groups targeted for recruitment.

Social media recruitment occurs in a public or semipublic sphere [63], which means that the distinction between private and public communication is not always as transparent as in traditional social settings. For instance, social media users might have their own perceptions of private and public spaces on social media [96]. Concurrently, users may not be aware of what purposes their communication is used for or even who can access the content they publish [63,97] (see the *Challenge: Digital Literacy* section). This potentially leads to private information being disclosed involuntarily, whereby users may unwillingly grant access to information on their health status to others or are unable to anticipate the potential consequences of publicly disclosed health-related information (see the *Challenge: Privacy and Data Security* section). In such situations, social media recruitment could lead to stigmatization [3,98]. Fear of stigmatization can be a significant barrier to participation in clinical studies [77] but occur only if individuals are identifiable from disease-related activities, groups, or comments [3]. The risk of stigmatization is closely related to privacy issues [99] and violates the principle of nonmaleficence [17,100]. As marginalized groups are at particular risk of stigmatization on social media, this stands in sharp contrast to the benefit of reaching marginalized groups, as outlined in the *Benefit: Including Marginalized Groups by Accessing Hard-to-Reach Populations* section.

Empirical studies have suggested that stigmatized health conditions are treated differently from nonstigmatized conditions on social media, and individuals with stigmatized diseases prefer anonymity when discussing their condition on the web [101-103]. In general, as Boudewyns et al [104] have shown using examples of sexually transmitted diseases, people talk less about the health conditions to which they attach a stigma. In this respect, interaction with clinical studies in the context of social media recruitment might involuntarily lift this anonymity. Conversely, other findings have suggested that patients with stigmatized conditions turn to social media to build relationships with other patients, acquire new information, and receive emotional support [105-108].

Challenge: Equality of Access

Although social media makes certain traditionally hard-to-reach populations more accessible for clinical study recruitment, it might also lower the chances of certain participants accessing the clinical study, particularly those underrepresented on social media. Consequently, the recruitment strategy must consider potential access issues, particularly when recruiting from closed groups [15]. Various empirical evidence suggests that despite the increasing spread of mobile devices [109,110], some groups are less likely than others to engage with technological devices and, therefore, have less access to social media. First, this may be relevant to older adults, as older age is associated with poor digital literacy [111-113] and lower internet use [114]. Second, individuals with lower educational attainment appear to be underrepresented on social media [15,115]. Moreover, low socioeconomic status tends to correlate with low eHealth literacy [116], and individuals with a low socioeconomic status tend to use the internet in *more general and superficial ways* [117,118] (see also the *Challenge: Digital Literacy* section). Thus, recruitment via social media may be challenging when these groups are targeted.

Although the ability of many social media platforms to direct study advertisements toward defined target groups can be useful in addressing unknown patients, it is unclear how the underlying algorithms choose members of the target group [13,119]. Although researchers choose certain patient parameters such as age or residential area, Facebook also uses its activity logs to improve the target algorithm [120]. Hence, researchers have only limited control over who is getting the advertisement and whether this remains stable over time. Consequently, such algorithms not only represent an ethical challenge related to privacy (see the *Challenge: Privacy and Data Security* section) but also make it difficult for researchers to control for the equitable distribution of information and advertisements, which provides challenges related to research quality (see the *Challenge: Research Quality* section). The magnitude of this issue depends on the social media platform and the strategy used for recruitment and, therefore, needs a context-specific assessment.

XSL•FO
RenderX

Beneficence

Benefit: Increasing Effectiveness and Cost-effectiveness of Recruitment

For ethical reasons, recruitment for clinical studies should be designed as effectively and cost-effectively as possible. Furthermore, choosing the most cost-effective recruitment option is key to achieving the maximum benefit from a given research budget. Early termination of studies because of a lack of participants puts existing study participants through pointless risks and could impair their trust in the research [121]. However, the lack of effective recruitment remains a persistent challenge in clinical studies.

According to some estimates, up to 60% of clinical trials are delayed or canceled because of a lack of enrollment [3-5]. However, social media recruitment strategies have shown early signs of effectiveness in various types of clinical trials [3], including studies of HIV vaccines [8], smoking cessation studies [6], studies of bothersome vulvovaginal symptoms [122], studies targeting patients with type 1 diabetes [7], occipital nerve studies [123], and studies of depression prevention [124]. A systematic review also reported shorter recruitment periods than traditional strategies in health research [88]. However, as success stories are typically published more often than failures, it is expected that the actual effectiveness depends on the target group and recruitment strategy.

Approximately half of the studies that recruited participants for medical research reported that social media is more cost-effective than traditional methods [9]. In a recent review, Brøgger-Mikkelsen et al [125] found that the median cost per enrollee for web-based recruitment strategies was US $72, and the median cost per enrollee for offline recruitment strategies was US $199, with 31% (4/13) of the included studies reporting web-based recruitment to be less cost-effective than offline recruitment. However, clinical trials provide a specific context as clinicians routinely come into contact with potential participants, and inclusion criteria are usually complex and involve clinical data [126,127].

Challenge: Research Quality

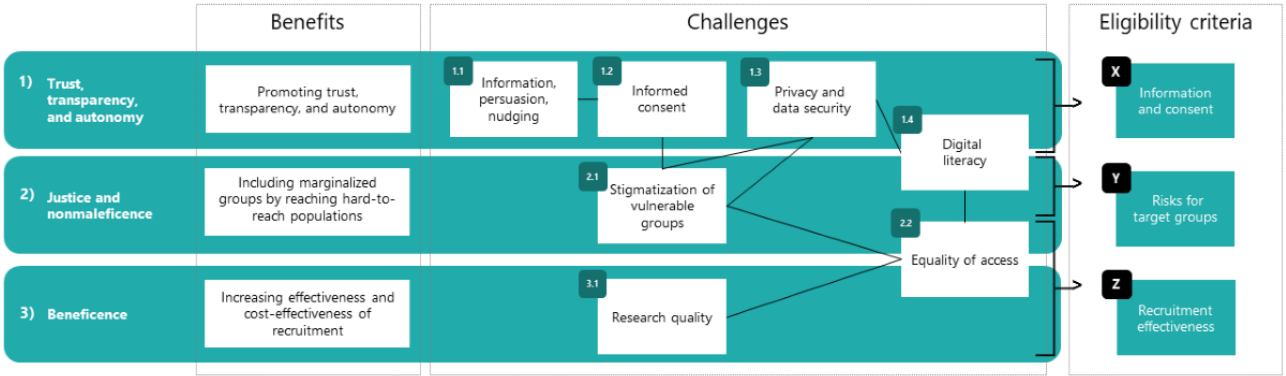
Existing evidence for recruitment effectiveness and cost-effectiveness of social media recruitment suggests that success is context dependent. The reasons for low recruitment accrual include the overall presentation of the study to the participant possibly being inappropriate or that the chosen platform does not adequately cover the demographic profile of the target population [49]. These issues, along with opaque social media algorithms (refer to the *Challenge: Equality of Access* section) and a digital literacy divide (refer to the *Challenge: Digital Literacy* section), could lead to the low statistical representativeness of those recruited via social media [15,109]. Overrepresented demographics typically include higher education, young age, and lack of immigration history [6,9,128], leading to ethnically homogeneous samples [129]. Although such issues of representativeness are a major concern for population-based studies, whether such issues are applicable depends on the target population of the clinical study. However, social media recruitment does not necessarily lead to effective recruitment results from either a qualitative or a quantitative perspective.

Part B: Eligibility Criteria of Ethical Social Media Recruitment

Overview

By reviewing the conceptual and empirical scientific literature, we identified and analyzed the benefits and challenges of social media recruitment for clinical studies. We found that the ethical and practical eligibility of social media recruitment needs to be assessed separately for each study, as ethical benefits and challenges are highly context sensitive. On the basis of the review of ethical issues (part A), we propose three criteria to assess the ethical and practical eligibility of social media recruitment (see Figure 1 for an overview): criterion X, *Information and Consent*; criterion Y, *Risks for Target Groups*; and criterion Z, *Recruitment Effectiveness*. In the following sections, we present the practical implications of each of these criteria and provide an example of how these criteria can be visualized.

Figure 1. Overview of the benefits and challenges of social media recruitment (part A) and the eligibility criteria to be used for context-specific assessments of social media recruitment strategies.



XSL•FO
RenderX

Criterion X: Information and Consent

Transparency

Several forms of transparency must be met in the context of clinical study recruitment on social media to address potential issues related to nudging, informed consent, and privacy.

Investigator transparency means that researchers make their identity as researchers visible when interacting with users on social media regarding a study [3].

Data transparency includes the disclosure of collected, stored, and used data in study-related social media activities. A major challenge is the nontransparent data processing on most social media platforms, which includes black box machine learning models to predict future behavior [64,65,67]. Open-source social media platforms are currently not widespread and therefore might not be suitable as an effective recruitment strategy. Consequently, researchers often rely on private sector tools for recruitment via social media, where they do not have complete control or are not aware of the potential use of the data collected during the recruitment process. At the very least, this should be acknowledged, and social media users should be made aware of it, particularly in the context of clinical studies.

Information transparency is necessary to obtain informed consent. As social media recruitment includes unique issues related to privacy and stigma (refer to the *Challenge: Privacy and Data Security* and *Challenge: Stigmatization of the Vulnerable* sections; also refer to the study by Gelinas et al [3]) that differ from ethical issues arising from the clinical study itself, we propose to distinguish between consent for social media recruitment and traditional patient consent. For social media recruitment, informed consent would depend on how the participants are contacted. Platforms such as Facebook offer several ways to access potential participants [130], such as paid advertisements, where target groups can be specified based on demographic characteristics, interests, and previous web-based activities; project-specific pages, where information about the study is posted on Facebook pages of existing groups related to the topic of interest; and by directly contacting potential participants via private messages [131]. In the following sections, we expand on the questions of whether it is necessary to obtain separate consent for social media recruitment and whether, and to what extent, researchers are responsible for increasing the digital literacy of social media users.

Is It Necessary to Obtain Informed Consent for the Recruitment Process on Social Media?

We argue that this depends on the recruitment strategy. Table 1 summarizes what we consider the 4 types of recruitment strategies that have implications for information provision and consent. For the type A and type B strategies, no separate consent is needed as no data are collected on the social media platform. However, it might be necessary to supply information about the study with a disclaimer alerting people of the potential implications concerning their privacy if they share information about the study with others. In contrast, type C and type D strategies include direct contact between researchers or research-related social media channels and (potential) research participants. Although this holds advantages in terms of patient engagement and trust building, it also opens up other issues as study-related data are collected on social media platforms. Participants should be informed about how these data will be used in the context of the study, with whom it will be shared, and how and when it will be deleted, and they should give explicit consent at the beginning of the interaction. The same applies if potential participants are actively recruited in private groups, where they expect to be in a private environment (eg, closed groups on Facebook). Some forms of consent, such as through the group moderator, for contacting these individuals on social media is an ethical imperative in these instances [3]. In type D recruitment strategies, researchers use participants' social networks to identify and actively address other potential research participants. Although this issue and its implications for privacy and consent have been discussed at length by Gelinas et al [3], we would like to add to their argument that such recruitment strategies should be avoided in particularly vulnerable target populations because of the increased risk of stigmatization, harmful privacy violations, and other psychosocial side effects [98,99].

Table 1. SMR[a] strategies and their implications for information and transparency.

SMR strategy type	Aim	Methods	Characteristics	Scope of targeted audience	Implications for informed consent
Type A	Raise awareness for the study	Advertisements and posts	No engagement in the study on social media	Targeting a broad audience	No signed consent needed
Type B	Actively include social media users in recruiting participants	Sharing of posts, advertisements, and informative material by users	Users are encouraged to reveal connections to a clinical study	Targeting a broad audience	No signed consent is needed, but a disclaimer raising awareness of disclosing connection to the study is required
Type C	Using closed groups for recruitment and community management	Dialogs between researchers and users or in between users; postings in private groups	Study-related data are collected on social media	Targeting a narrow audience	Explicit consent required
Type D	Using the user's social networks to identify potential participants	Private messaging; active network research of users	Potential user information is revealed to others; user data collected for the study	Targeting a narrow audience	Explicit consent required; caution with vulnerable groups

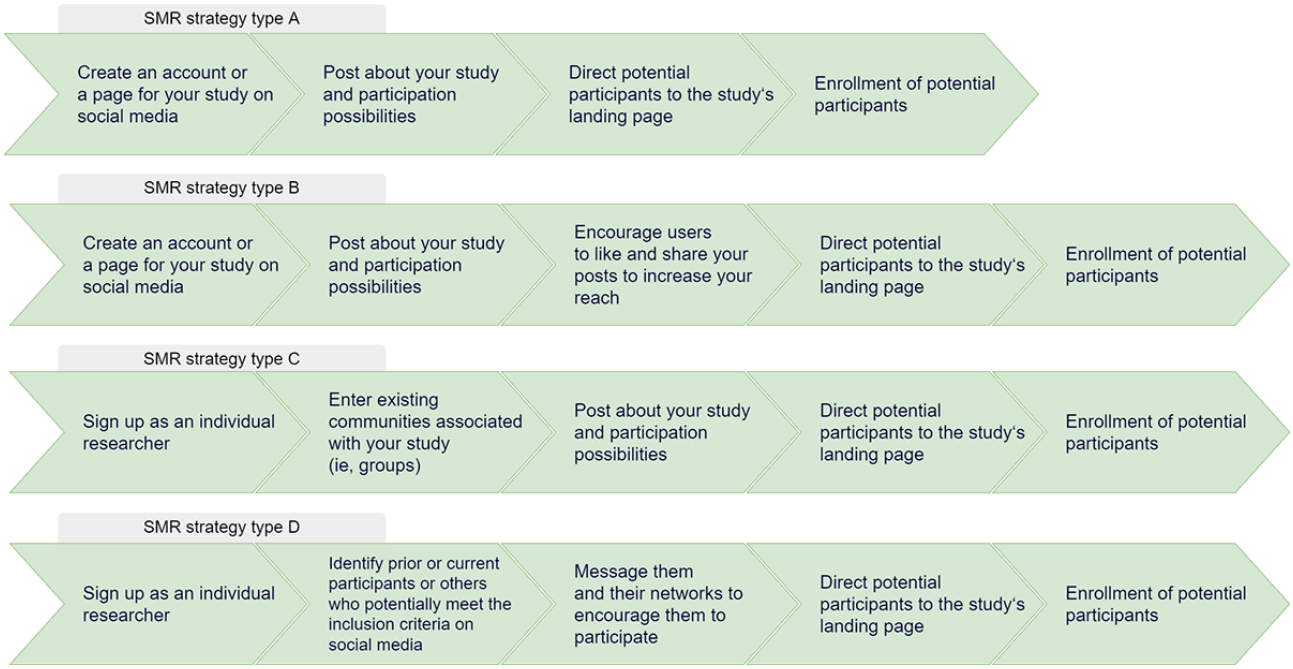
[a]SMR: social media recruitment.

XSL•FO
RenderX

Are Researchers Recruiting for Clinical Studies on Social Media Responsible for Improving the Digital Literacy of Social Media Users?

A lack of digital literacy (refer to the *Challenge: Digital Literacy* section) is connected to other ethical challenges, including privacy and stigmatization issues and inadequate patient information [31,71,72,109,110]. On the basis of previous findings [73-76], researchers engaging in social media–based recruitment for clinical studies cannot assume that potential participants are aware of these issues. Researchers are responsible for providing sufficient information to ensure that potential participants do not harm themselves because of a lack of knowledge or awareness. In practice, when using social media recruitment only to raise awareness of the existence of a study (type A or B; Table 1), a small disclaimer may be sufficient. For community engagement in the context of a clinical study (type C or D; Table 1), we recommend developing codes of conduct in community groups and information materials in the form of quizzes or small videos. Such materials should include information on how to protect other patients and provide awareness of the potential risks (including privacy violations and stigma) attached to social media use. Figure 2 illustrates the practical differences of these recruitment types.

Figure 2. Mock recruitment strategies to illustrate recruitment types A to D. SMR: social media recruitment.



Criterion Y: Risks for Target Groups

Overview

As outlined previously, it appears that both using social media and *not* using it could potentially lead to discrimination and stigmatization. Therefore, the vulnerabilities of target groups should be carefully assessed for web-based and offline recruitment. Considering the broad definition of vulnerability as the incapability to protect one's interests [19], we propose that the digital literacy of the target group and its risk to attract social stigma as well as the social media recruitment type should be taken into account.

Digital Literacy of the Target Group

The literature indicates that insufficient digital literacy is associated with low socioeconomic and educational status and older age [116]. Therefore, target groups meeting one or several of these criteria should be considered vulnerable to social media recruitment. They may be unable to access clinical studies recruiting only on the web and be more prone to misunderstanding the context of clinical studies in the web-based environment. In addition, young target groups tend to underestimate privacy issues on social media and should therefore be contacted with particular care [74,75].

Social Stigma of Disease or Other Characteristics

Diseases or characteristics with a social or structural stigma attached to them are varied and can depend on the sociocultural context. These may include characteristics related to sexual orientation [132,133], sexually transmitted diseases [134-136], psychiatric disorders [137,138], or skin diseases [139,140]. Target groups should be carefully evaluated in terms of stigmatized characteristics (see Multimedia Appendix 2 for details), and protective privacy measures should be intensified accordingly. For example, if social media recruitment leads to the public outing of individuals experiencing stigmatized diseases, privacy is particularly important.

Recruitment Type

Recruiting from existing participant networks (type D strategies; Table 1) holds particular risks for target groups and should only be considered after explicit consent is given and if the target group is not considered vulnerable, as outlined previously.

Criterion Z: Recruitment Effectiveness

Overview

To ensure good research quality and effectiveness and avoid unintended harm, researchers should avoid adding social media recruitment as an explorative and inexpensive alternative to

XSL•FO
RenderX

other recruitment methods without a detailed implementation plan. Instead, the added value of using social media should be critically examined in the context of a specific clinical study, and a recruitment strategy should be planned a priori when applying for funding. Several dimensions should be considered.

Target Group Definition

When considering social media for clinical study recruitment, the age and socioeconomic distribution of the target population must be considered to ensure research quality and improve equality of access. Social media recruitment should only be used if the target group is available on these platforms.

Platform Choice

Depending on the technical features, user numbers, user groups, policies, prices, and other characteristics, some social media platforms might be more appropriate for clinical study recruitment than others. These can be used differently by the recruiter by building upon the different features offered by the platforms. For instance, age distribution and other user characteristics vary considerably across platforms [7,49,115]. Therefore, it might be helpful to use multi-platform approaches and triangulate them with other recruitment strategies to avoid inequalities stemming from populations that are inactive on social media or those with low digital literacy [131].

Anticipating Patient Responses

Depending on the study design, inclusion criteria, target groups, and recruitment strategy, responses of potential participants will be more or less numerous and accurate in terms of eligibility criteria. For example, type A and type B recruitment strategies target a broader audience, and more (and nonspecific) responses might be expected than in type C and type D strategies (Table 1). The extent of patient responses must be aligned with the resources available to respond to them, particularly when targeting patients with chronic or severe diseases and/or very specific inclusion criteria. For these patients, it might be harmful if they were placing hope in a study they had heard about on social media but then never received a response. However, it might be overwhelming and frustrating for the personnel responsible for patient recruitment if many of the responding patients are not eligible for the study or if resources are not sufficient to respond to all requests. The extent of this problem depends on the type of clinical study.

Expertise Within the Research Team

High-quality planning and execution of social media recruitment require specialized knowledge from the research team. Therefore, the clinical study staff should receive formal training, and dedicated recruiters for social media should be employed. Interdisciplinary skills such as platform-specific expertise in terms of use, science communication, and illustrative skills should be required and either represented in the research team or provided through an external provider.

The Eligibility Matrix

Depending on the context of a clinical study, the 3 eligibility criteria might have different weights for the overall assessment of social media recruitment. We suggest weighing the risks for target groups (criterion Y) as particularly high as this criterion directly corresponds to the principle of nonmaleficence [141,142]. This criterion usually cannot be improved by an adapted social media recruitment strategy, as it depends on the study-specific target population. In contrast, the other two criteria (X and Z) can be addressed in the recruitment strategy.

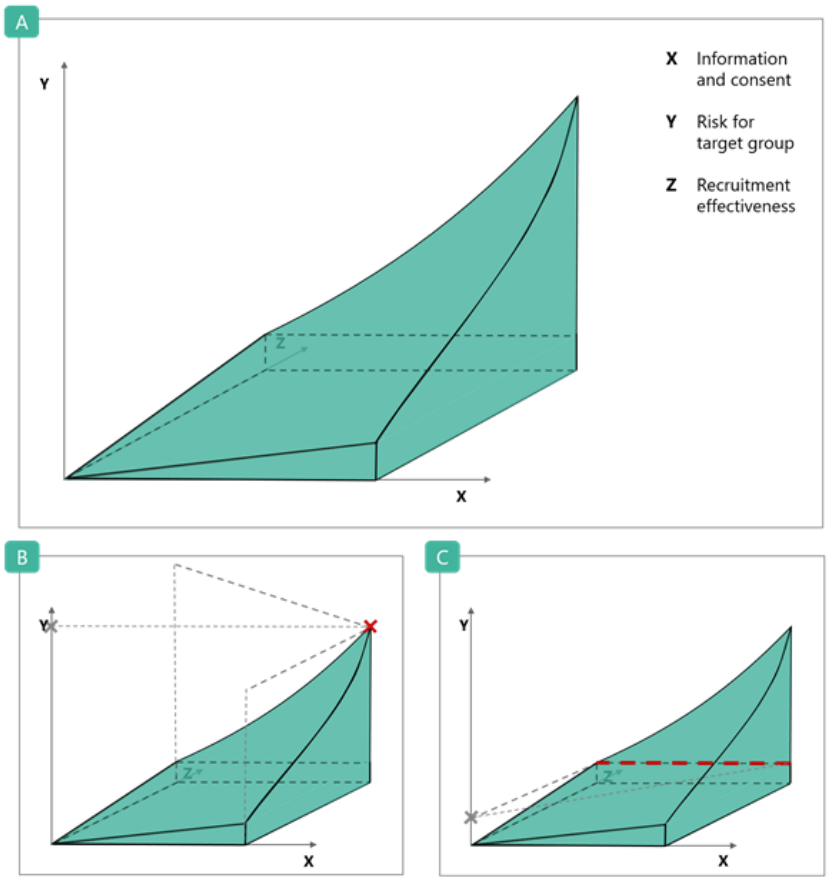
To visualize how the 3 eligibility criteria depend on each other, we arranged them in a 3D matrix (Figures 3A-3C). In cases where there are substantial risks for the target group, social media recruitment would only be permissible in very specific cases (eg, if it is not possible to recruit in any other way or if alternative recruitment methods pose even greater risks). Figures 3B and 3C present 2 examples of these dependencies.

First, if the estimated risk for the target group (y-axis) is particularly high, the estimated recruitment effectiveness (z-axis) also needs to be high, and the informed consent procedure (x-axis) must be extensive for social media recruitment to be considered in a study (Figure 3B). For instance, if we want to recruit participants who have been diagnosed with a stigmatized disease, such as HIV or hepatitis B, the recruitment risks for the target population (Y) are substantial (refer to the *Challenge: Stigmatization of the Vulnerable* section). In this case, social media recruitment must be both effective (X) and have a thoroughly informed consent procedure (Z) to serve as an appropriate recruitment method.

Second, if the overall risk for the target population (y-axis) is estimated to be low and the estimated effectiveness (z-axis) is very high, the range for an acceptable informed consent procedure (x-axis) would broaden (Figure 3C). For example, if we aim to recruit for a clinical study focusing on common health-related behaviors such as vitamin intake and exercise or menopausal symptoms, the risks of recruitment (Y) are lower for the target population. If we expect moderate to good effectiveness in recruiting through social media (X) as we are, for instance, explicitly interested in young participants, the informed consent procedure (Z) only needs to meet the minimum requirements to be adequate.

XSL•FO
RenderX

Figure 3. (A) Eligibility matrix for the assessment of social media recruitment for clinical studies. If the result of an assessment of the 3 dimensions occurs within the green volume, social media recruitment can be considered ethical for a particular study. Suitability for using social media for recruitment increases with the respective increase in dimensions X and Z (X: informed consent and Z: recruitment effectiveness) and decreases with a rising risk for the target group (Y: risk for target group). (B) Dashed lines indicate an example of a high-risk target group. X indicates a very limited scope for social media recruitment. (C) Dashed lines indicate an example of a low-risk target group. Dashes represent the scope of the informed consent procedure.



Discussion

Principal Findings

Social media recruitment can offer important benefits for clinical studies, including reaching hard-to-reach populations; increasing recruitment effectiveness; and supporting trust, transparency, and autonomy. These potential advantages make it an ethical imperative to consider this recruitment strategy as a supplement to traditional strategies. However, potential risks strongly depend on the study design, target population, and details of the recruitment strategy. Therefore, each strategy should take into account the target population and the potential risks they might face and should be approved by an institutional review board or ethics committee. In the context of clinical studies, it might be worthwhile to limit social media contact with patients to avoid enhancing the risks of privacy violations or stigmatization and for researchers to create a thorough risk assessment and details on how the recruitment strategy will take these risks into account. However, as social media interactions also potentially increase trust, transparency, and participant engagement, these risks should be weighted context specifically.

If a clinical study targets a particularly vulnerable population, a solution could be to target multiplicators via social media, such as social workers or general practitioners, who then inform the eligible patients about the study. This would lower the risk of privacy violations and stigmatization in vulnerable groups and provide a more beneficial risk-benefit assessment. A potential disadvantage of this approach concerns the indirect steps taken toward recruitment; that is, success depends on the activity of the multiplicator. However, in clinical studies, patients have to come into direct contact with the health care system, and the recruitment process is not as straightforward as it would be for web-based studies.

We identified several research gaps related to the potentially trust-building features of social media recruitment through active engagement, informed consent on the platform itself, and risks of aggravated stigmatization. Relatedly, scholars have identified a lack of ethical and regulatory guidelines, as well as missing reporting standards, that ensure transparency related to social media recruitment for clinical studies [143]. Collecting empirical evidence on the perceptions of researchers, users, and patients concerning social media recruitment for clinical studies is a prerequisite for developing such guidelines. Furthermore, it should be acknowledged that such guidelines are best applied as context specific, as existing privacy and data protection regulations differ between different regions of the world.

Some large and well-known social media platforms (eg, Facebook and Twitter) are more frequently used for clinical study recruitment than others. Thus, they are better documented in empirical studies. As our analysis was partly based on empirical cases of clinical studies that used social media

XSL•FO
RenderX

platforms as recruitment tools, other social media platforms might have additional ethical implications for recruitment that were not covered in this contribution.

Conclusions

Ethical challenges related to social media recruitment are context sensitive. We suggest that the most important challenges for social media recruitment can be assessed by evaluating three dimensions: the level of information and consent, risks for target groups, and effectiveness of the recruitment strategy. These dimensions are interconnected and should be evaluated strategically, critically, and repeatedly. In Multimedia Appendix 2, we provide a checklist with practical recommendations for clinical researchers considering social media recruitment.

Social media recruitment for clinical studies is becoming increasingly common and should only be approved and executed if planned and assessed appropriately. This is particularly important in clinical studies, which might come with additional ethical implications. We suggest that researchers designing a clinical study should use the matrix we have presented to assess a priori whether they should use social media for recruitment.

Acknowledgments

This work was supported by the European Union's Horizon 2020 research and innovation program (grant agreement no. 848223).The authors thank Dr Nina Goldman for her critical revisions and language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search algorithms for literature review.
[PDF File (Adobe PDF File), 76 KB - jmir_v24i5e31231_app1.pdf]

Multimedia Appendix 2
Checklist for researchers considering social media recruitment for clinical studies.
[PDF File (Adobe PDF File), 493 KB - jmir_v24i5e31231_app2.pdf]

References

1. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. Adv psychiatr treat 2018 Jan 02;9(3):229-238. [doi: 10.1192/apt.9.3.229]
2. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14(1):42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
3. Gelinas L, Pierce R, Winkler S, Cohen IG, Lynch HF, Bierer BE. Using social media as a research recruitment tool: ethical issues and recommendations. Am J Bioeth 2017 Mar;17(3):3-14 [FREE Full text] [doi: 10.1080/15265161.2016.1276644] [Medline: 28207365]
4. Puffer S, Torgerson DJ. Recruitment difficulties in randomised controlled trials. Controlled Clin Trial 2003 Dec;24:214S-215S [FREE Full text] [doi: 10.1016/s0197-2456(03)00148-x]
5. Wertheimer A. Non-completion and informed consent. J Med Ethics 2014 Feb;40(2):127-130. [doi: 10.1136/medethics-2012-101108] [Medline: 23371314]
6. Frandsen M, Thow M, Ferguson SG. The effectiveness of social media (Facebook) compared with more traditional advertising methods for recruiting eligible participants to health research studies: a randomized, controlled clinical trial. JMIR Res Protoc 2016 Aug 10;5(3):e161 [FREE Full text] [doi: 10.2196/resprot.5747] [Medline: 27511829]
7. Wisk LE, Nelson EB, Magane KM, Weitzman ER. Clinical trial recruitment and retention of college students with type 1 diabetes via social media: an implementation case study. J Diabetes Sci Technol 2019 Apr 22;13(3):445-456 [FREE Full text] [doi: 10.1177/1932296819839503] [Medline: 31010315]
8. Sitar S, Hartman B, Graham B, Ledgerwood J. P15-17. Social media as a tool for engaging and educating audiences around HIV vaccine research and clinical trial participation. Retrovirology 2009 Oct 22;6:P218. [doi: 10.1186/1742-4690-6-s3-p218]
9. Topolovec-Vranic J, Natarajan K. The use of social media in recruitment for medical research studies: a scoping review. J Med Internet Res 2016 Nov 07;18(11):e286 [FREE Full text] [doi: 10.2196/jmir.5698] [Medline: 27821383]
10. Balthazar P, Harri P, Prater A, Safdar NM. Protecting your patients' interests in the era of big data, artificial intelligence, and predictive analytics. J Am Coll Radiol 2018 Mar;15(3 Pt B):580-586. [doi: 10.1016/j.jacr.2017.11.035] [Medline: 29402532]
11. Ghani NA, Hamid S, Targio Hashem IA, Ahmed E. Social media big data analytics: a survey. Comput Human Behav 2019 Dec;101:417-428. [doi: 10.1016/j.chb.2018.08.039]

XSL•FO
RenderX

12. Lassen NB, Cour LL, Vatrapu R. Predictive analytics with social media data. In: Sloan L, Quan-Haase A, editors. The SAGE Handbook of Social Media Research Methods. Thousand Oaks, California: SAGE Publications; 2017.
13. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
14. Lutz H, Schiebel M, Tuider E, editors. Handbuch Biographieforschung. Wiesbaden: Springer; 2018.
15. Kamp K, Herbell K, Magginis WH, Berry D, Given B. Facebook recruitment and the protection of human subjects. West J Nurs Res 2019 Sep 07;41(9):1270-1281 [FREE Full text] [doi: 10.1177/0193945919828108] [Medline: 30729866]
16. van Dijck J. The Culture of Connectivity A Critical History of Social Media. Oxfordshire, England, UK: Oxford University Press; 2013.
17. Beauchamp T, Childress J. Principles of Biomedical Ethics 7th edition. Oxfordshire, England, UK: Oxford University Press; 2013.
18. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
19. International Ethical Guidelines for Health-Related Research Involving Humans. Geneva, Switzerland: CIOMS; 2016.
20. Thompson MA, O'Regan RM. Social media and clinical trials: the pros and cons gain context when the patient is at the center. Cancer 2018 Dec 15;124(24):4618-4621 [FREE Full text] [doi: 10.1002/cncr.31747] [Medline: 30339288]
21. Giustini D. Social media and clinical trials recruitment: potential benefits and challenges. J Can Health Libr Assoc 2014 Jul 22;33(3):140. [doi: 10.5596/c12-044]
22. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, et al. Ethical issues of social media usage in healthcare. Yearb Med Inform 2015 Aug 13;10(1):137-147 [FREE Full text] [doi: 10.15265/IY-2015-001] [Medline: 26293861]
23. Close S, Smaldone A, Fennoy I, Reame N, Grey M. Using information technology and social networking for recruitment of research participants: experience from an exploratory study of pediatric Klinefelter syndrome. J Med Internet Res 2013 Mar 19;15(3):e48 [FREE Full text] [doi: 10.2196/jmir.2286] [Medline: 23512442]
24. Thompson MA. Social media in clinical trials. Am Soc Clin Oncol Educ Book 2014 May(34):e101-e105. [doi: 10.14694/edbook_am.2014.34.e101]
25. Fiske A, Prainsack B, Buyx A. Meeting the needs of underserved populations: setting the agenda for more inclusive citizen science of medicine. J Med Ethics 2019 Sep 12;45(9):617-622 [FREE Full text] [doi: 10.1136/medethics-2018-105253] [Medline: 31300487]
26. Fiske A, Del Savio L, Prainsack B, Buyx A. Conceptual and ethical considerations for citizen science in biomedicine. In: Heyen N, Dickel S, Brüninghaus A, editors. Personal Health Science: Persönliches Gesundheitswissen zwischen Selbstsorge und Bürgerforschung. Wiesbaden: Springer VS; 2019:195-217.
27. Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences: opportunities, challenges, ethical considerations, and practical guidelines. Am Psychol 2015 Sep;70(6):543-556. [doi: 10.1037/a0039210] [Medline: 26348336]
28. Fiske A, Prainsack B, Buyx A. Data work: meaning-making in the era of data-rich medicine. J Med Internet Res 2019 Jul 09;21(7):e11672 [FREE Full text] [doi: 10.2196/11672] [Medline: 31290397]
29. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
30. Bleilevens W, Schenk J. Das digitale patientenzentrierte Pharmaunternehmen. In: Digitale Transformation von Dienstleistungen im Gesundheitswesen VII: Impulse für die Pharmaindustrie. Wiesbaden: Springer Gabler; 2020:1-16.
31. O'Connor A, Jackson L, Goldsmith L, Skirton H. Can I get a retweet please? Health research recruitment and the Twittersphere. J Adv Nurs 2014 Mar;70(3):599-609. [doi: 10.1111/jan.12222] [Medline: 23909740]
32. Martinez O, Wu E, Shultz AZ, Capote J, López Rios J, Sandfort T, et al. Still a hard-to-reach population? Using social media to recruit Latino gay couples for an HIV intervention adaptation study. J Med Internet Res 2014 Apr 24;16(4):e113 [FREE Full text] [doi: 10.2196/jmir.3311] [Medline: 24763130]
33. MacInnis DJ, Jaworski BJ. Information processing from advertisements: toward an integrative framework. J Mark 2018 Nov 02;53(4):1-23. [doi: 10.1177/002224298905300401]
34. Thaler R, Sunstein C. Nudge: Improving Decisions About Health, Wealth, and Happiness. New Haven, CT: Penguin; 2008.
35. Prainsack B. The value of healthcare data: to nudge, or not? Policy Stud 2020 Feb 04;41(5):547-562. [doi: 10.1080/01442872.2020.1723517]
36. van den Hoven M. Nudging for others' sake: an ethical analysis of the legitimacy of nudging healthcare workers to accept influenza immunization. Bioethics 2021 Feb 13;35(2):143-150 [FREE Full text] [doi: 10.1111/bioe.12819] [Medline: 33051889]
37. European Observatory on Health Systems and Policies, Prainsack B, Buyx A. Nudging and solidarity: do they go together? Eurohealth 2014;20(2):14-17.
38. Engelen B. Ethical criteria for health-promoting nudges: a case-by-case analysis. Am J Bioeth 2019 May;19(5):48-59. [doi: 10.1080/15265161.2019.1588411] [Medline: 31068115]

XSL•FO
RenderX

39. Cohen S. Nudging and informed consent. Am J Bioeth 2013 Jun;13(6):3-11. [doi: 10.1080/15265161.2013.781704] [Medline: 23641835]
40. Gelfand SD. The meta-nudge - a response to the claim that the use of nudges during the informed consent process is unavoidable. Bioethics 2016 Oct 13;30(8):601-608. [doi: 10.1111/bioe.12266] [Medline: 27411164]
41. Kroll T, Stieglitz S. Digital nudging and privacy: improving decisions about self-disclosure in social networks. Behav Inf Technol 2019 Feb 27;40(1):1-19 [FREE Full text] [doi: 10.1080/0144929x.2019.1584644]
42. Puaschunder JM. Nudging in the digital big data era. Eur J Econ 2017 Dec 31;04(04):18-23. [doi: 10.19044/elp.v4no4a2]
43. Wisniewski PJ, Knijnenburg BP, Lipford HR. Making privacy personal: profiling social network users to inform privacy education and nudging. Int J Human Comput Stud 2017 Feb;98:95-108. [doi: 10.1016/j.ijhcs.2016.09.006]
44. VanEpps EM, Volpp KG, Halpern SD. A nudge toward participation: improving clinical trial enrollment with behavioral economics. Sci Transl Med 2016 Jul 20;8(348):348fs13 [FREE Full text] [doi: 10.1126/scitranslmed.aaf0946] [Medline: 27440726]
45. Krutsinger DC, O'Leary KL, Ellenberg SS, Cotner CE, Halpern SD, Courtright KR. A randomized controlled trial of behavioral nudges to improve enrollment in critical care trials. Annals ATS 2020 Sep;17(9):1117-1125. [doi: 10.1513/annalsats.202003-194oc]
46. Blumenthal-Barby JS, Burroughs H. Seeking better health care outcomes: the ethics of using the "nudge". Am J Bioeth 2012 Feb;12(2):1-10. [doi: 10.1080/15265161.2011.634481] [Medline: 22304506]
47. Dunn LB, Gordon NE. Improving informed consent and enhancing recruitment for research by understanding economic behavior. JAMA 2005 Feb 02;293(5):609-612. [doi: 10.1001/jama.293.5.609] [Medline: 15687316]
48. Korn L, Betsch C, Böhm R, Meier NW. Social nudging: the effect of social feedback interventions on vaccine uptake. Health Psychol 2018 Nov;37(11):1045-1054. [doi: 10.1037/hea0000668] [Medline: 30221969]
49. Arigo D, Pagoto S, Carter-Harris L, Lillie SE, Nebeker C. Using social media for health research: methodological and ethical considerations for recruitment and intervention delivery. Digit Health 2018 May 07;4:2055207618771757 [FREE Full text] [doi: 10.1177/2055207618771757] [Medline: 29942634]
50. Böhme R, Köpsell S. Trained to accept?: a field experiment on consent dialogs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI '10: CHI Conference on Human Factors in Computing Systems; Apr 10 - 15, 2010; Atlanta Georgia USA. [doi: 10.1145/1753326.1753689]
51. Ramo DE, Hall SM, Prochaska JJ. Reaching young adult smokers through the internet: comparison of three recruitment mechanisms. Nicotine Tob Res 2010 Jul;12(7):768-775 [FREE Full text] [doi: 10.1093/ntr/ntq086] [Medline: 20530194]
52. Wozney L, Turner K, Rose-Davis B, McGrath PJ. Facebook ads to the rescue? Recruiting a hard to reach population into an internet-based behavioral health intervention trial. Internet Interv 2019 Sep;17:100246 [FREE Full text] [doi: 10.1016/j.invent.2019.100246] [Medline: 31080751]
53. Adam LM, Manca DP, Bell RC. Can Facebook be used for research? Experiences using Facebook to recruit pregnant women for a randomized controlled trial. J Med Internet Res 2016 Sep 21;18(9):e250 [FREE Full text] [doi: 10.2196/jmir.6404] [Medline: 27655184]
54. Jones R, Lacroix LJ, Porcher E. Facebook advertising to recruit young, urban women into an HIV prevention clinical trial. AIDS Behav 2017 Nov;21(11):3141-3153 [FREE Full text] [doi: 10.1007/s10461-017-1797-3] [Medline: 28528463]
55. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/ajph.2017.304249]
56. Rothwell E, Wong B, Rose NC, Anderson R, Fedor B, Stark LA, et al. A randomized controlled trial of an electronic informed consent process. J Empir Res Hum Res Ethics 2014 Dec 02;9(5):1-7 [FREE Full text] [doi: 10.1177/1556264614552627] [Medline: 25747685]
57. Lunt H, Connor S, Skinner H, Brogden G. Electronic informed consent: the need to redesign the consent process for the digital age. Intern Med J 2019 Jul 11;49(7):923-929. [doi: 10.1111/imj.14339] [Medline: 31295775]
58. Welch BM, Marshall E, Qanungo S, Aziz A, Laken M, Lenert L, et al. Teleconsent: a novel approach to obtain informed consent for research. Contemp Clin Trials Commun 2016 Aug 15;3:74-79 [FREE Full text] [doi: 10.1016/j.conctc.2016.03.002] [Medline: 27822565]
59. Krishnamurti T, Argo N. A patient-centered approach to informed consent. Med Decis Making 2016 Mar 10;36(6):726-740. [doi: 10.1177/0272989x16636844]
60. Smith JH, Dinev T, Xu H. Information privacy research: an interdisciplinary review. MIS Q 2011;35(4):989. [doi: 10.2307/41409970]
61. Solove DJ. A taxonomy of privacy. University Pennsylvania Law Rev 2006;154(3):477-564. [doi: 10.2307/40041279]
62. Bélanger F, Crossler R. Privacy in the digital age: a review of information privacy research in information systems. MIS Q 2011;35(4):1017. [doi: 10.2307/41409971]
63. McKee R. Ethical issues in using social media for health and health care research. Health Policy 2013 May;110(2-3):298-301. [doi: 10.1016/j.healthpol.2013.02.006] [Medline: 23477806]
64. Thorson K, Cotter K, Medeiros M, Pak C. Algorithmic inference, political interest, and exposure to news and politics on Facebook. Inf Commun Soc 2019 Jul 27;24(2):183-200. [doi: 10.1080/1369118x.2019.1642934]

XSL•FO
RenderX

65. Hazelwood K, Bird S, Brooks D, Chintala S, Diril U, Dzhulgakov D, et al. Applied machine learning at Facebook: a datacenter infrastructure perspective. In: Proceedings of the 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA). 2018 Presented at: 2018 IEEE International Symposium on High Performance Computer Architecture (HPCA); Feb 24-28, 2018; Vienna, Austria. [doi: 10.1109/hpca.2018.00059]
66. O'Neil C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. London, United Kingdom: Penguin Books Limited; 2016.
67. Mühlhoff R. Predictive privacy: towards an applied ethics of data analytics. Ethics Inf Technol 2021 Jul 31;23(4):675-690. [doi: 10.1007/s10676-021-09606-x]
68. Wachter S, Mittelstadt B. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum Bus L Rev 2019;494:499-505. [doi: 10.31228/osf.io/mu2kf]
69. Eshet Y. Digital literacy: a conceptual framework for survival skills in the digital era. J Educ Multimedia Hypermedia 2004;13(1):106 [FREE Full text]
70. Gilster P. Digital Literacy. Hoboken, New Jersey: Wiley; 1997.
71. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
72. Frampton GK, Shepherd J, Pickett K, Griffiths G, Wyatt JC. Digital tools for the recruitment and retention of participants in randomised controlled trials: a systematic map. Trials 2020 Jun 05;21(1):478 [FREE Full text] [doi: 10.1186/s13063-020-04358-3] [Medline: 32498690]
73. Boyd D. Making sense of privacy and publicity. SXSW. 2010. URL: http://www.danah.org/papers/talks/2010/SXSW2010.html [accessed 2022-04-03]
74. Madden M, Lenhart A, Cortesi S, Gasser U, Duggan M, Smith A, et al. Teens, social media, and privacy. Pew Research Center. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2013/05/PIP_TeensSocialMediaand Privacy_PDF.pdf [accessed 2022-04-03]
75. Parsi K, Elster N. Conducting research on social media--is Facebook like the public square? Am J Bioeth 2014 Sep 17;14(10):63-65. [doi: 10.1080/15265161.2014.947825] [Medline: 25229596]
76. Rivera-Romero O, Konstantinidis S, Denecke K, Gabarrón E, Petersen C, Househ M, et al. Ethical considerations for participatory health through social media: healthcare workforce and policy maker perspectives. Yearb Med Inform 2020 Aug 17;29(1):71-76 [FREE Full text] [doi: 10.1055/s-0040-1701981] [Medline: 32303101]
77. Clark LT, Watkins L, Piña IL, Elmer M, Akinboboye O, Gorham M, et al. Increasing diversity in clinical trials: overcoming critical barriers. Curr Probl Cardiol 2019 May;44(5):148-172 [FREE Full text] [doi: 10.1016/j.cpcardiol.2018.11.002] [Medline: 30545650]
78. Hussain-Gambles M, Atkin K, Leese B. Why ethnic minority groups are under-represented in clinical trials: a review of the literature. Health Soc Care Community 2004 Sep;12(5):382-388. [doi: 10.1111/j.1365-2524.2004.00507.x] [Medline: 15373816]
79. Braunholtz DA, Edwards SJ, Lilford RJ. Are randomized clinical trials good for us (in the short term)? Evidence for a "trial effect". J Clin Epidemiol 2001 Mar;54(3):217-224. [doi: 10.1016/s0895-4356(00)00305-x]
80. Michaels JA. Potential for epistemic injustice in evidence-based healthcare policy and guidance. J Med Ethics 2020 May 27;47(6):417-422. [doi: 10.1136/medethics-2020-106171] [Medline: 32461243]
81. Caplan A, Friesen P. Health disparities and clinical trial recruitment: is there a duty to tweet? PLoS Biol 2017 Mar;15(3):e2002040 [FREE Full text] [doi: 10.1371/journal.pbio.2002040] [Medline: 28249024]
82. Ejiogu N, Norbeck JH, Mason MA, Cromwell BC, Zonderman AB, Evans MK. Recruitment and retention strategies for minority or poor clinical research participants: lessons from the Healthy Aging in Neighborhoods of Diversity across the Life Span study. Gerontologist 2011 Jun;51 Suppl 1:S33-S45 [FREE Full text] [doi: 10.1093/geront/gnr027] [Medline: 21565817]
83. Carlini BH, Safioti L, Rue TC, Miles L. Using Internet to recruit immigrants with language and culture barriers for tobacco and alcohol use screening: a study among Brazilians. J Immigr Minor Health 2015 Apr;17(2):553-560. [doi: 10.1007/s10903-013-9934-1] [Medline: 24563138]
84. Lohse B. Facebook is an effective strategy to recruit low-income women to online nutrition education. J Nutr Educ Behav 2013;45(1):69-76. [doi: 10.1016/j.jneb.2012.06.006] [Medline: 23305805]
85. Burton-Chase AM, Parker WM, Hennig K, Sisson F, Bruzzone LL. The use of social media to recruit participants with rare conditions: lynch syndrome as an example. JMIR Res Protoc 2017 Jan 23;6(1):e12 [FREE Full text] [doi: 10.2196/resprot.6066] [Medline: 28115298]
86. Schaffer A, Cairney J, Cheung A, Veldhuizen S, Levitt A. Community survey of bipolar disorder in Canada: lifetime prevalence and illness characteristics. Can J Psychiatry 2006 Jan 01;51(1):9-16. [doi: 10.1177/070674370605100104] [Medline: 16491979]
87. King DB, O'Rourke N, DeLongis A. Social media recruitment and online data collection: a beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations. Canadian Psychology/Psychologie canadienne 2014;55(4):240-249. [doi: 10.1037/a0038087]

XSL•FO
RenderX

88. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]
89. Fitzpatrick T, Zhou K, Cheng Y, Chan P, Cui F, Tang W, et al. A crowdsourced intervention to promote hepatitis B and C testing among men who have sex with men in China: study protocol for a nationwide online randomized controlled trial. BMC Infect Dis 2018 Sep 29;18(1):489 [FREE Full text] [doi: 10.1186/s12879-018-3403-3] [Medline: 30268114]
90. Levine C, Faden R, Grady C, Hammerschmidt D, Eckenwiler L, Sugarman J, Consortium to Examine Clinical Research Ethics. The limitations of "vulnerability" as a protection for human research participants. Am J Bioeth 2004 Aug;4(3):44-49. [doi: 10.1080/15265160490497083] [Medline: 16192138]
91. Hurst S. Vulnerability in research and health care; describing the elephant in the room? Bioethics 2008 May;22(4):191-202. [doi: 10.1111/j.1467-8519.2008.00631.x] [Medline: 18405317]
92. Research ethics, vulnerability, and trust on the internet. In: Second International Handbook of Internet Research. Dordrecht: Springer; 2018.
93. Racine E, Bracken-Roche D. Enriching the concept of vulnerability in research ethics: an integrative and functional account. Bioethics 2019 Jan 23;33(1):19-34. [doi: 10.1111/bioe.12471] [Medline: 30136737]
94. Lange MM, Rogers W, Dodds S. Vulnerability in research ethics: a way forward. Bioethics 2013 Jul 30;27(6):333-340. [doi: 10.1111/bioe.12032] [Medline: 23718774]
95. Cascio MA, Racine E. Person-oriented research ethics: integrating relational and everyday ethics in research. Account Res 2018 Mar 02;25(3):170-197. [doi: 10.1080/08989621.2018.1442218] [Medline: 29451025]
96. Eysenbach G, Till JE. Ethical issues in qualitative research on internet communities. BMJ 2001 Nov 10;323(7321):1103-1105 [FREE Full text] [doi: 10.1136/bmj.323.7321.1103] [Medline: 11701577]
97. Hartz T, Fangerau H, Albrecht UV. Social Media in der ärztlichen Praxis: chancen, risiken, trends. Deutsches Ärzteblatt 2014;111(17 Suppl. PRAXIS):6 [FREE Full text]
98. Bender JL, Cyr AB, Arbuckle L, Ferris LE. Ethics and privacy implications of using the internet and social media to recruit participants for health research: a privacy-by-design framework for online recruitment. J Med Internet Res 2017 Apr 06;19(4):e104 [FREE Full text] [doi: 10.2196/jmir.7029] [Medline: 28385682]
99. Akers L, Gordon JS. Using Facebook for large-scale online randomized clinical trial recruitment: effective advertising strategies. J Med Internet Res 2018 Nov 08;20(11):e290 [FREE Full text] [doi: 10.2196/jmir.9372] [Medline: 30409765]
100. Côté I. Fieldwork in the era of social media: opportunities and challenges. PS Political Sci Politics 2013;46(03):615-619. [doi: 10.1017/s1049096513000565]
101. Rains SA. The implications of stigma and anonymity for self-disclosure in health blogs. Health Commun 2014;29(1):23-31. [doi: 10.1080/10410236.2012.714861] [Medline: 23356432]
102. de Choudhury M, Morris MR, White RW. Seeking and sharing health information online: comparing search engines and social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Apr 26 - May 1, 2014; Toronto Ontario Canada. [doi: 10.1145/2556288.2557214]
103. Pavalanathan U, de Choudhury M. Identity management and mental health discourse in social media. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW '15: 24th International World Wide Web Conference; May 18 - 22, 2015; Florence Italy. [doi: 10.1145/2740908.2743049]
104. Boudewyns V, Himelboim I, Hansen DL, Southwell BG. Stigma's effect on social interaction and social media activity. J Health Commun 2015;20(11):1337-1345. [doi: 10.1080/10810730.2015.1018604] [Medline: 26087307]
105. Andalibi N. What happens after disclosing stigmatized experiences on identified social media: individual, dyadic, and social/network outcomes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland Uk. [doi: 10.1145/3290605.3300367]
106. Highton-Williamson E, Priebe S, Giacco D. Online social networking in people with psychosis: a systematic review. Int J Soc Psychiatry 2015 Feb;61(1):92-101. [doi: 10.1177/0020764014556392] [Medline: 25381145]
107. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
108. Sannon S, Murnane EL, Bazarova NN, Gay G. "I was really, really nervous posting it": communicating about Invisible Chronic Illnesses across Social Media Platforms. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland Uk. [doi: 10.1145/3290605.3300583]
109. Oesterle S, Epstein M, Haggerty KP, Moreno MA. Using Facebook to recruit parents to participate in a family program to prevent teen drug use. Prev Sci 2018 May;19(4):559-569 [FREE Full text] [doi: 10.1007/s11121-017-0844-7] [Medline: 29116552]
110. Technology device ownership: 2015. Pew Research Center. URL: https://www.pewresearch.org/internet/2015/10/29/technology-device-ownership-2015/ [accessed 2022-04-03]
111. 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 Nov 27;11(4):e48 [FREE Full text] [doi: 10.2196/jmir.1249] [Medline: 19945947]

XSL•FO
RenderX

112. Prensky M. Digital natives, digital immigrants part 1. On Horizon 2001 Sep;9(5):1-6. [doi: 10.1108/10748120110424816]
113. Czaja SJ, Charness N, Fisk AD, Hertzog C, Nair SN, Rogers WA, et al. Factors predicting the use of technology: findings from the center for research and education on aging and technology enhancement (CREATE). Psychol Aging 2006 Jun;21(2):333-352 [FREE Full text] [doi: 10.1037/0882-7974.21.2.333] [Medline: 16768579]
114. Friemel TN. The digital divide has grown old: determinants of a digital divide among seniors. New Media Soc 2014 Jun 12;18(2):313-331. [doi: 10.1177/1461444814538648]
115. Social media use in 2018. Pew Research Center. URL: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ [accessed 2022-04-03]
116. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
117. van Deursen AJ, van Dijk JA. The digital divide shifts to differences in usage. New Media Soc 2013 Jun 07;16(3):507-526. [doi: 10.1177/1461444813487959]
118. The Deepening Divide: Inequality in the Information Society. Thousand Oaks, California: SAGE Publications; 2005.
119. Facebook algorithms and personal data. Pew Research Center. URL: https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/ [accessed 2022-04-03]
120. Eslami S, Kumaran SR, Sandvig C, Karahalios K. Communicating algorithmic process in online behavioral advertising. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: CHI Conference on Human Factors in Computing Systems; Apr 21 - 26, 2018; Montreal QC Canada. [doi: 10.1145/3173574.3174006]
121. Malmqvist E, Juth N, Lynöe N, Helgesson G. Early stopping of clinical trials: charting the ethical terrain. Kennedy Inst Ethics J 2011 Mar;21(1):51-78. [doi: 10.1353/ken.2011.0002] [Medline: 21598846]
122. Guthrie KA, Caan B, Diem S, Ensrud KE, Greaves SR, Larson JC, et al. Facebook advertising for recruitment of midlife women with bothersome vaginal symptoms: a pilot study. Clin Trials 2019 Oct;16(5):476-480 [FREE Full text] [doi: 10.1177/1740774519846862] [Medline: 31055949]
123. Goadsby PJ. Analysis of occipital nerve stimulation in studies of chronic migraine and broader implications of social media in clinical trials. Cephalalgia 2013 Feb;33(3):214-215. [doi: 10.1177/0333102412468680] [Medline: 23212293]
124. Morgan AJ, Jorm AF, Mackinnon AJ. Internet-based recruitment to a depression prevention intervention: lessons from the Mood Memos study. J Med Internet Res 2013 Feb 12;15(2):e31 [FREE Full text] [doi: 10.2196/jmir.2262] [Medline: 23403043]
125. Brøgger-Mikkelsen M, Ali Z, Zibert JR, Andersen AD, Thomsen SF. Online patient recruitment in clinical trials: systematic review and meta-analysis. J Med Internet Res 2020 Nov 04;22(11):e22179 [FREE Full text] [doi: 10.2196/22179] [Medline: 33146627]
126. Fletcher B, Gheorghe A, Moore D, Wilson S, Damery S. Improving the recruitment activity of clinicians in randomised controlled trials: a systematic review. BMJ Open 2012 Jan 06;2(1):e000496 [FREE Full text] [doi: 10.1136/bmjopen-2011-000496] [Medline: 22228729]
127. Penberthy L, Brown R, Puma F, Dahman B. Automated matching software for clinical trials eligibility: measuring efficiency and flexibility. Contemp Clin Trials 2010 May;31(3):207-217 [FREE Full text] [doi: 10.1016/j.cct.2010.03.005] [Medline: 20230913]
128. Koo M, Skinner H. Challenges of internet recruitment: a case study with disappointing results. J Med Internet Res 2005 Mar 19;7(1):e6 [FREE Full text] [doi: 10.2196/jmir.7.1.e6] [Medline: 15829478]
129. Layi G, Albright CA, Berenberg J, Plant K, Ritter P, Laurent D, et al. UH Cancer Center Hotline: recruiting cancer survivors for an online health-behavior change intervention: are different strategies more beneficial? Hawaii Med J 2011 Oct;70(10):222-223 [FREE Full text] [Medline: 22162599]
130. Amon KL, Campbell AJ, Hawke C, Steinbeck K. Facebook as a recruitment tool for adolescent health research: a systematic review. Acad Pediatr 2014;14(5):439-447. [doi: 10.1016/j.acap.2014.05.049] [Medline: 25169155]
131. Reuter K. Social media for clinical trial recruitment: how real is the potential? EMJ Innov 2020 Jan 9:34-39 [FREE Full text] [doi: 10.33590/emjinnov/19-00121]
132. Pachankis JE, Hatzenbuehler ML, Berg RC, Fernández-Dávila P, Mirandola M, Marcus U, et al. Anti-LGBT and anti-immigrant structural stigma: an intersectional analysis of sexual minority men's HIV risk when migrating to or within Europe. J Acquir Immune Defic Syndr 2017 Dec 01;76(4):356-366 [FREE Full text] [doi: 10.1097/QAI.0000000000001519] [Medline: 28787329]
133. Whitehead J, Shaver J, Stephenson R. Outness, stigma, and primary health care utilization among rural LGBT populations. PLoS One 2016 Jan 5;11(1):e0146139 [FREE Full text] [doi: 10.1371/journal.pone.0146139] [Medline: 26731405]
134. Valizadeh L, Zamanzadeh V, Bayani M, Zabihi A. The social stigma experience in patients with hepatitis B infection. Gastroenterol Nursing 2017;40(2):143-150. [doi: 10.1097/sga.0000000000000223]
135. Chan RC, Mak WW, Ma GY, Cheung M. Interpersonal and intrapersonal manifestations of HIV stigma and their impacts on psychological distress and life satisfaction among people living with HIV: toward a dual-process model. Qual Life Res 2021 Jan 09;30(1):145-156. [doi: 10.1007/s11136-020-02618-y] [Medline: 32909160]

XSL•FO
RenderX

136. Fitzgerald-Husek A, Van Wert MJ, Ewing WF, Grosso AL, Holland CE, Katterl R, et al. Measuring stigma affecting sex workers (SW) and men who have sex with men (MSM): a systematic review. PLoS One 2017 Nov 30;12(11):e0188393 [FREE Full text] [doi: 10.1371/journal.pone.0188393] [Medline: 29190642]
137. Miller PK, Cuthbertson CA, Loveridge S. Social status influence on stigma towards mental illness and substance use disorder in the United States. Community Ment Health J 2022 Feb 04;58(2):249-260. [doi: 10.1007/s10597-021-00817-6] [Medline: 33817761]
138. Nohr L, Lorenzo Ruiz A, Sandoval Ferrer JE, Buhlmann U. Mental health stigma and professional help-seeking attitudes a comparison between Cuba and Germany. PLoS One 2021 Feb 11;16(2):e0246501 [FREE Full text] [doi: 10.1371/journal.pone.0246501] [Medline: 33571232]
139. Wu JH, Cohen BA. The stigma of skin disease. Curr Opinion Pediatrics 2019;31(4):509-514. [doi: 10.1097/mop.0000000000000792]
140. Dimitrov D, Szepietowski JC. Instruments to assess stigmatization in dermatology. Postepy Hig Med Dosw (Online) 2017 Nov 03;71:901-905 [FREE Full text] [doi: 10.5604/01.3001.0010.5607] [Medline: 29151062]
141. Page K. The four principles: can they be measured and do they predict ethical decision making? BMC Med Ethics 2012 May 20;13(1):10 [FREE Full text] [doi: 10.1186/1472-6939-13-10] [Medline: 22606995]
142. Landau R, Osmo R. Professional and personal hierarchies of ethical principles. Int J Soc Welfare 2003;12(1):42-49. [doi: 10.1111/1468-2397.00007]
143. Reuter K, Zhu Y, Angyan P, Le N, Merchant AA, Zimmer M. Public concern about monitoring Twitter users and their conversations to recruit for clinical trials: survey study. J Med Internet Res 2019 Oct 30;21(10):e15455 [FREE Full text] [doi: 10.2196/15455] [Medline: 31670698]

Edited by R Kukafka; submitted 14.06.21; peer-reviewed by R Pierce, Y Zhang, K Devon; comments to author 30.07.21; revised version received 11.11.21; accepted 02.12.21; published 03.05.22.

Please cite as:
Zimmermann BM, Willem T, Bredthauer CJ, Buyx A
Ethical Issues in Social Media Recruitment for Clinical Studies: Ethical Analysis and Framework
J Med Internet Res 2022;24(5):e31231
URL: https://www.jmir.org/2022/5/e31231
doi:10.2196/31231
PMID:35503247

©Bettina M Zimmermann, Theresa Willem, Carl Justus Bredthauer, Alena Buyx. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2022. 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.

XSL•FO
RenderX

Original Paper

Distinguishing Admissions Specifically for COVID-19 From Incidental SARS-CoV-2 Admissions: National Retrospective Electronic Health Record Study

Jeffrey G Klann[1], MEng, PhD; Zachary H Strasser[1], MD; Meghan R Hutch[2], BS; Chris J Kennedy[3], PhD; Jayson S Marwaha[4], MD; Michele Morris[5], BA; Malarkodi Jebathilagam Samayamuthu[5], MD; Ashley C Pfaff[4], MD; Hossein Estiri[1], PhD; Andrew M South[6], MD, MS; Griffin M Weber[7], MD, PhD; William Yuan[7], PhD; Paul Avillach[7], MD, PhD; Kavishwar B Wagholikar[1], MBBS, PhD; Yuan Luo[2], PhD; The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)[7]; Gilbert S Omenn[8], MD, PhD; Shyam Visweswaran[5], MD, PhD; John H Holmes[9], MS, PhD; Zongqi Xia[10], MD, PhD; Gabriel A Brat[7], MD; Shawn N Murphy[11], MD, PhD

[1]Laboratory of Computer Science, Department of Medicine, Massachusetts General Hospital, Boston, MA, United States

[2]Department of Preventive Medicine, Northwestern University, Chicago, IL, United States

[3]Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[4]Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States

[5]Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, United States

[6]Section of Nephrology, Department of Pediatrics, Brenner Children's, Wake Forest School of Medicine, Winston Salem, NC, United States

[7]see Acknowledgments

[8]Center for Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, United States

[9]Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[10]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States

[11]Department of Neurology, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Jeffrey G Klann, MEng, PhD
Laboratory of Computer Science
Department of Medicine
Massachusetts General Hospital
50 Staniford St
Suite 750
Boston, MA, 02114
United States
Phone: 1 6142822222
Email: jeff.klann@mgh.harvard.edu

Abstract

Background: Admissions are generally classified as COVID-19 hospitalizations if the patient has a positive SARS-CoV-2 polymerase chain reaction (PCR) test. However, because 35% of SARS-CoV-2 infections are asymptomatic, patients admitted for unrelated indications with an incidentally positive test could be misclassified as a COVID-19 hospitalization. Electronic health record (EHR)–based studies have been unable to distinguish between a hospitalization specifically for COVID-19 versus an incidental SARS-CoV-2 hospitalization. Although the need to improve classification of COVID-19 versus incidental SARS-CoV-2 is well understood, the magnitude of the problems has only been characterized in small, single-center studies. Furthermore, there have been no peer-reviewed studies evaluating methods for improving classification.

Objective: The aims of this study are to, first, quantify the frequency of incidental hospitalizations over the first 15 months of the pandemic in multiple hospital systems in the United States and, second, to apply electronic phenotyping techniques to automatically improve COVID-19 hospitalization classification.

XSL•FO
RenderX

Methods: From a retrospective EHR-based cohort in 4 US health care systems in Massachusetts, Pennsylvania, and Illinois, a random sample of 1123 SARS-CoV-2 PCR-positive patients hospitalized from March 2020 to August 2021 was manually chart-reviewed and classified as "admitted with COVID-19" (incidental) versus specifically admitted for COVID-19 ("for COVID-19"). EHR-based phenotyping was used to find feature sets to filter out incidental admissions.

Results: EHR-based phenotyped feature sets filtered out incidental admissions, which occurred in an average of 26% of hospitalizations (although this varied widely over time, from 0% to 75%). The top site-specific feature sets had 79%-99% specificity with 62%-75% sensitivity, while the best-performing across-site feature sets had 71%-94% specificity with 69%-81% sensitivity.

Conclusions: A large proportion of SARS-CoV-2 PCR-positive admissions were incidental. Straightforward EHR-based phenotypes differentiated admissions, which is important to assure accurate public health reporting and research.

(J Med Internet Res 2022;24(5):e37931) doi:10.2196/37931

KEYWORDS

COVID-19; medical informatics; public health; phenotype; electronic health records; clinical research informatics; health data; SARS-CoV-2; patient data; health care

Introduction

Despite the ongoing COVID-19 pandemic and the dozens of research groups and consortia worldwide that continue to utilize clinical data available in electronic health records (EHRs), critical gaps remain in both our understanding of COVID-19 and how to accurately predict poor outcomes, including hospitalization and mortality [1-4].

One of the most prominent gaps in the field is how to distinguish hospital admissions specifically for COVID-19-related indications (eg, severe disease with respiratory failure) from an incidentally positive SARS-CoV-2 polymerase chain reaction (PCR) test in admissions for an unrelated reason (eg, a broken leg). Approximately 800,000 new SARS-CoV-2 cases are being reported daily, and approximately 150,000 patients are hospitalized with a positive SARS-CoV-2 PCR test [5]. Misclassification of incidental COVID-19 during hospitalizations is common [5] and raises research and public health concerns. For example, deleterious effects on health care system resource disbursement or utilization as well as on local and regional social and economic structure and function can result from inaccurate reporting of incidental cases of SARS-CoV-2.

Misclassification in research studies occurs because patients are usually considered COVID-19 patients if they have a recent positive SARS-CoV-2 PCR test or the *International Classification of Diseases, Tenth Revision* (ICD-10) diagnosis code U07, which, according to guidelines, is equivalent to a positive test [6]. This approach has been used in most COVID-19 studies published to date [7,8] and is in line with Centers of Disease Control and Prevention (CDC) guidelines, which treat positive SARS-CoV-2 PCR tests as confirmed cases [9]. Given that at least 35% of SARS-CoV-2 cases are asymptomatic, patients seeking unrelated care are erroneously classified as COVID hospitalizations [10-14]. The magnitude of this misclassification has increased over time as health care systems began to be less restrictive after the second wave and elective surgeries were again performed starting in the second quarter of 2021.

A potential solution is EHR-based phenotyping, which identifies patient populations of interest based on proxies derived from EHR observations. EHR phenotypes are developed by first performing manual chart review to classify cases and then applying a machine learning or statistical reasoning method to the EHR data to create an explainable predictive model [15,16]. For example, a phenotyping study of bipolar disorder found that true bipolar disorder is correlated with a set of several EHR features [17]. Our previous work validated a "severe COVID-19" phenotype in the Consortium for Clinical Characterization of COVID-19 by EHR (4CE) network using both chart review and comparison across sites [18,19]. 4CE is a diverse international network of over 300 hospitals engaged in collaborative COVID-19 research [2,20,21].

The Massachusetts Department of Public Health has recently begun using a simple phenotype to report COVID-19 hospitalizations [22,23]. Although it is based on treatment recommendations and not a gold standard, it illustrates the interest in EHR-based phenotyping for COVID-19.

In this study, we utilized EHR data from 60 hospitals across 4 US health care systems in 4CE, combined with clinical expertise, data analytics, and manual EHR chart review, to determine whether patients admitted to the hospital and who had a positive SARS-CoV-2 PCR test were hospitalized for COVID-19 (for-COVID-19 group) or were admitted for a different indication and simply had an incidental positive test (admitted-with-COVID-19 group).

Methods

Sites

We selected a sample of 4 4CE sites across the United States to participate in the development of our for-COVID-19 hospitalization phenotype. These sites included the Beth Israel Deaconess Medical Center (BIDMC), Mass General Brigham (MGB), Northwestern University (NWU), and the University of Pittsburgh/University of Pittsburgh Medical Center (UPITT). Each site involved at least 1 *clinical* expert (for chart review and manual annotation) and 1 *data analytics* expert (to apply various analytic filtering approaches). Eligible patients for this study were those included in the 4CE COVID-19 cohort: all hospitalized patients (pediatric and adult) with their first positive SARS-CoV-2 PCR test 7 days before to 14 days after hospitalization [2].

XSL•FO
RenderX

Chart Review

Each development site randomly sampled an equal number of admissions in each quarter (BIDMC, MGB) or month (NWU, UPITT) from their cohort of SARS-CoV-2 PCR-positive patients over the period of March 2020 until at least March 2021 (N=1123). Clinical experts reviewed the charts in the EHRs and recorded whether these patients were admitted for COVID-19-related reasons, as defined later. The total number of chart reviews per site was somewhat variable and determined by availability of the clinical experts. Participating sites and the number of chart reviews are listed in Table 1.

To develop chart review criteria, a 4CE subgroup met during March-July 2021. The group consists of about 20 researchers in 4CE, with a mixture of physicians, medical informaticians, and data scientists. In the process, dozens of real patient charts were considered, and edge cases were discussed until consensus was reached on the minimal chart review necessary to determine the reason a patient was hospitalized.

Based on the developed criteria, chart reviewers (1 per site, except at the BIDMC, where there were 2) classified the patients based on review of primarily the admission note, discharge summary (or death note), and laboratory values for the hospitalization. Each site had Institutional Review Board (IRB) approval to view the charts locally, and only deidentified aggregate summaries were presented to the subgroup. Each site summarized the chart reviews in a spreadsheet that was then linked to the site's 4CE EHR data, wherein medical record numbers were replaced with 4CE's patient pseudoidentifiers, and criteria classifications were coded as an integer. The 4CE EHR data set is a COVID-19-related subset of raw EHR data consisting of selected laboratory test, medication, and procedure categories and all available ICD-10 diagnosis codes. The data dictionary is explained in more detail in Multimedia Appendix 1. The chart review process is presented visually in steps 1-5 of Figure 1.

We developed an R script (R Core Team) at the MGB to perform basic data summarization. This did the following: calculated chart review summary statistics, aggregated data on ICD-10 diagnosis codes used during the hospitalization to compare to the chart review classification, and generated a bubble plot that visualizes the change in proportion of hospitalizations, specifically for COVID-19, among all chart reviews over the course of the pandemic, by month. A trendline was fitted with locally estimated scatterplot smoothing (loess) regression using ggplot2 and was weighted by the number of chart reviews performed that month. Each participating health care site ran the R script on its chart-reviewed patient cohort.

Table 1. Participating health care systems' overall characteristics and the number and period of chart reviews performed for this study.

Participating site	Hospitals, n	Inpatient discharges per year, n	Number of chart reviews performed, n	Chart review time period, start date-end date
BIDMC[a]	1	40,752	400	March 2020-March 2021
MGB[b]	10	163,521	406	March 2020-July 2021
NWU[c]	10	103,279	70	March 2020-February 2021
UPITT[d]	39	369,300	247	April 2020-August 2021

[a]BIDMC: Beth Israel Deaconess Medical Center.

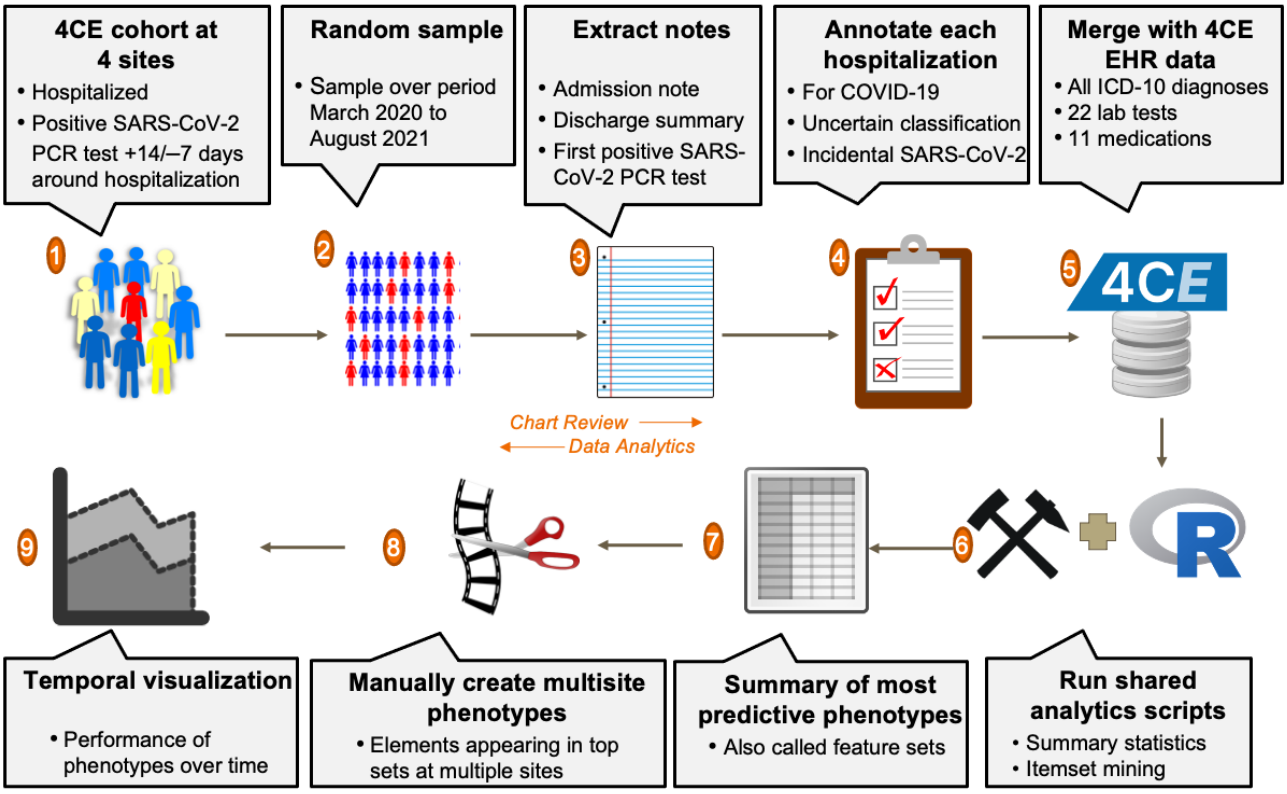
[b]MGB: Mass General Brigham.

[c]NWU: Northwestern University.

[d]UPITT: University of Pittsburgh/University of Pittsburgh Medical Center.

XSL•FO
RenderX

Figure 1. The chart review process. (1-2) At each site, an equal number of patients admitted with a positive SARS-CoV-2 PCR test were sampled by quarter or by month. (3-4) A chart reviewer at the site examined primarily the admission note, discharge summary (or death note), and laboratory values for the hospitalization to classify as admitted for COVID-19, incidental SARS-CoV2, or uncertain. (5-6) These classifications were then merged with 4CE EHR data for use with shared analytic scripts in R. (7-8) The top phenotypes at each site output by the data mining algorithm were summarized, and this was used to manually construct feature sets to be used across sites by selecting components that appeared in step 7 at multiple sites. (9) The performance over time of the top multisite phenotypes was visualized. 4CE: Consortium for Clinical Characterization of COVID-19 by EHR; EHR: electronic health record; ICD-10: International Classification of Diseases, Tenth Revision; PCR: polymerase chain reaction.



Phenotypes Using Hospital System Dynamics Phenotyping

We developed an algorithm as an R script to choose phenotypes of admissions specifically for COVID-19, using established hospital dynamics measures of ordering/charting patterns in the EHRs (eg, presence of laboratory tests rather than laboratory results) [16,24]. The algorithm uses a variation of the Apriori item set–mining algorithm [25,26]. Apriori, which has been utilized in other EHR studies, uses a hill-climbing approach to find iteratively larger item sets that meet some summary statistic constraint [27,28]. Apriori, like other market basket analyses, is advantageous when the labeled data are small, because it discovers statistical properties of the underlying data, rather than developing a separate predictive model that must be evaluated. Therefore, it does not require a data split between a training and a test set, which would further limit the sample size. The original algorithm chose rules that maximized the positive predictive value (PPV) and had at least a minimum prevalence in the data set. More recent variants use other summary statistics [29] because the PPV, which measures the likelihood a positive is a true positive, is highly affected by population prevalence (which shifts dramatically over time with COVID-19). Therefore, our algorithm used sensitivity and specificity. A visual representation of our algorithm is shown in Figure 2. Item sets of size 1 are chosen that meet certain minimum prediction thresholds, and then these are combined into item sets of size 2 and again filtered by the thresholds, and so forth up to a maximum item set size.

We applied our algorithm to find patterns in 4CE EHR data at each site using the presence of medications, laboratory tests, and diagnoses to select the best phenotypes. (Laboratory test results are included in the 4CE data set but were not included in this analysis, because it does not fit with the principles of hospital system dynamics [HSD].) We further compared the output at each site to see whether there were similarities (eg, transfer learning was applicable). We considered 2 cases: data that would be available in near-real time during a hospitalization (laboratory tests) and data that would be available for a retrospective analysis (including laboratory and medication facts and diagnosis codes, which are usually not coded until after discharge).

Sites exported phenotypes with sensitivity of at least 0.60, ordered by specificity in descending order. (Site B applied a slightly lower sensitivity threshold because no phenotypes with sensitivity of at least 0.60 were available.) Specificity was chosen as the sorting variable because it measures the phenotype's ability to detect and remove incidental SARS-CoV-2 admissions—a good measure of overall performance. Sensitivity, in contrast, measures the ability to select for-COVID-19 admissions, which can be easily maximized by simply selecting all patients. Groups of phenotypes were manually summarized into conjunctive normal

XSL•FO
RenderX

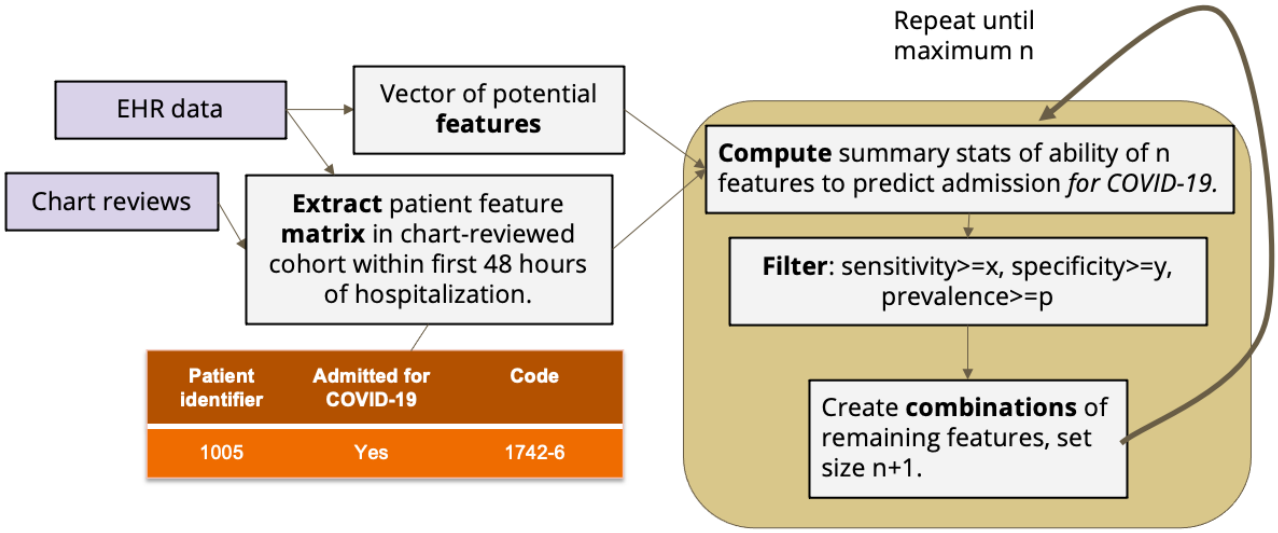
form by combining AND and OR phenotypes at each site, when possible, and reporting a sensitivity and specificity range for the final combined phenotype. We excluded feature sets that were more complex but with the same performance as a simpler feature set.

We also ran our phenotyping program to find the most predictive individual features at each site during every 6-month period of the pandemic, beginning January 2020. This analysis allowed us to examine the trend of HSD as the pandemic progressed.

The final piece of analysis involved selecting multisite phenotypes and plotting their performance over time. First, we selected the features that appeared at multiple sites from the best phenotypes at each site. We used these to manually construct multisite phenotypes. We optimized these using MGB data by manually adding/removing OR components based on performance, because adding too many OR components degrades the specificity. We ran these constructed phenotypes at each site to ascertain their performance characteristics.

This data-mining process can be seen visually in Figure 1, steps 6-8.

Figure 2. Design of the phenotyping algorithm. Predictive feature sets of iteratively larger size were selected based on their sensitivity and specificity in correctly identifying COVID-19-specific admissions using 4CE EHR data and chart reviews. We chose the following parameters after testing various thresholds at all 4 sites: AND feature sets, x=0.40, y=0.20, p=0.30; OR feature sets x=0.10, y=0.50, p=0.20; and single features: x=y=p=0. 4CE: Consortium for Clinical Characterization of COVID-19 by EHR; EHR: electronic health record.



Temporal Visualization of Phenotypes

We also developed a temporal visualization used at each site (step 9 of Figure 1). The visualization shows 3 lines: a solid line showing the total number of patients in the site's 4CE cohort (ie, admitted with a positive SARS-CoV-2 PCR test), a dashed line showing the total number of those patients after filtering to select patients admitted specifically for COVID-19 (ie, removing all patients who do not meet the phenotyping feature set criteria), and a dotted line showing the difference between the solid line and the dashed line (ie, patients removed from the cohort in the dashed line). Dots on the graph visualize the performance on the chart-reviewed cohort. Green dots on each line show patients who were correctly classified by the phenotype, according to the chart review. Likewise, orange dots on each line show incorrect classifications. Dot size is proportional to the number of chart reviews.

Importantly, all review and analysis were performed by local experts at each site, and only the final aggregated results were submitted to a central location for finalization. This approach is a hallmark of 4CE—keeping data close to local experts and only sharing aggregated results. It reduces regulatory complexity around data sharing and keeps those who know the data best involved in the analysis.

All our software tools were implemented as R programs. They were developed at the MGB and tested by all 4 sites. The code is available as open source [30].

Ethical Considerations

IRB approval was obtained at the BIDMC (#2020P000565), the MGB (#2020P001483), the UPITT (STUDY20070095), and the NWU (STU00212845). Participant informed consent was waived by each IRB because the study involved only retrospective data and no individually identifiable data were share outside of each site's local study team. Site names were anonymized (to sites A, B, C, and D) to comply with hospital privacy policies. At the MGB and the BIDMC, any counts of patients were blurred with a random number +/–3 before being shared centrally. Our previous work shows that for large counts, pooling blurred counts has minimal impact on the overall accuracy of the statistics [31]. At all sites, any counts <3 were censored. All other statistics (eg, percentages, differences, CIs, *P* values) were preserved.

Results

Chart Review

The final chart review criteria are shown in Table 2. (See the Methods section for details.) Across the 4 sites, 764 (68%) of

XSL•FO
RenderX

1123 patients were admitted *for* COVID-19, 292 (26%) patients were admitted with *incidental* SARS-CoV-2, and 67 (6%) were uncertain (Table 3). The 4 sites included the BIDMC, the MGB, the UPITT, and the NWU. A site-by-site breakdown, both overall and by individual criteria, is also shown in Table 3. A demographic characterization of the chart-reviewed cohort at each site is shown in Table 4. Plots of the proportion of hospitalizations specifically for COVID-19 among all chart reviews by month over the course of the pandemic are shown in Figure 3. Finally, Tables 5 and 6 show the top 10 ICD-10 diagnoses that were assigned to patients with a date in the first 48 hours after admission in *for*-COVID-19 versus *incidental*-COVID-19 groups. In all results, each site is labeled with a random but consistent letter (A, B, C, or D) to comply with hospital privacy policies.

Table 2. Summary of the chart review criteria developed by the 4CE[a] subgroup of physicians, medical informaticians, and data scientists.

Chart-reviewed classification	Criteria
Admitted specifically *for* COVID-19	Symptoms on admission *were attributable* to COVID-19, and clinicians admitted patients for COVID-19-related care. The symptoms included: • Respiratory insufficiency • Blood clots in vital organs • Hemodynamic changes • Other common viral symptoms, such as cough and fever • Admitted for non-COVID-19 issue but developed any of the above symptoms while hospitalized
Admitted incidentally *with* COVID-19	The admission history was *unlikely* to be related to COVID-19, and clinicians did not specifically admit the patient for COVID-19-related care. This admission could be due to: • Trauma • Procedure or operation requiring hospitalization • Term labor • Alternative causes, including drug overdose, cancer progression, and nonrespiratory severe infection
Uncertain	Symptoms on admission *may have been* related to COVID-19, and clinicians considered COVID-19 exacerbation during hospitalization. The symptoms included: • Preterm labor • Liver dysfunction • Graft failure • Immune system dysfunction • Alternative causes, including sickle cell crisis, failure to thrive, and altered mental status

[a]4CE: Consortium for Clinical Characterization of COVID-19 by EHR[b].

[b]EHR: electronic health record.

XSL•FO
RenderX

Table 3. Proportion of chart-reviewed patients admitted specifically for COVID-19 vs admitted with incidental SARS-CoV-2, overall and stratified by site, with a detailed criteria breakdown. A detailed breakdown at site D could not be included, because their process did not record the specific criteria for each classification. Note that cells with 0% are still included to show all the chart review criteria.

Category	Site A (N=406), n (%)	Site B (N=70), n (%)	Site C (N=247), n (%)	Site D (N=400), n (%)	Overall (N=1123), n (%)
Admitted specifically *for* COVID-19					764 (68)
All	288 (71)	59 (84)	180 (73)	240 (60)	N/A[a]
Respiratory insufficiency	202 (50)	36 (51)	128 (52)	N/A	N/A
Blood clot	6 (1)	<3 (<5)	<3 (<5)	N/A	N/A
Hemodynamic changes	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Other symptomatic COVID-19	71 (18)	19 (27)	47 (20)	N/A	N/A
Not admitted for COVID-19 but developed 1 of the above criteria	8 (2)	<3 (<5)	5 (2)	N/A	N/A
Admitted incidentally with COVID-19					292 (26)
All	85 (20)	9 (13)	54 (22)	144 (36)	N/A
Full-term labor	18 (4)	<3 (<5)	<3 (<5)	N/A	N/A
Procedure	8 (2)	<3 (<5)	9 (4)	N/A	N/A
Trauma	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Other not COVID-19	50 (13)	6 (9)	44 (18)	N/A	N/A
Uncertain					67 (6)
All	33 (8)	<3 (<5)	10 (4)	16 (4)	N/A
Immune dysfunction	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Early labor	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Liver dysfunction	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Graft failure	<3 (<5)	<3 (<5)	<3 (<5)	N/A	N/A
Other possible COVID-19	31 (8)	<3 (<5)	10 (4)	N/A	N/A

[a]N/A: not applicable.

XSL•FO
RenderX

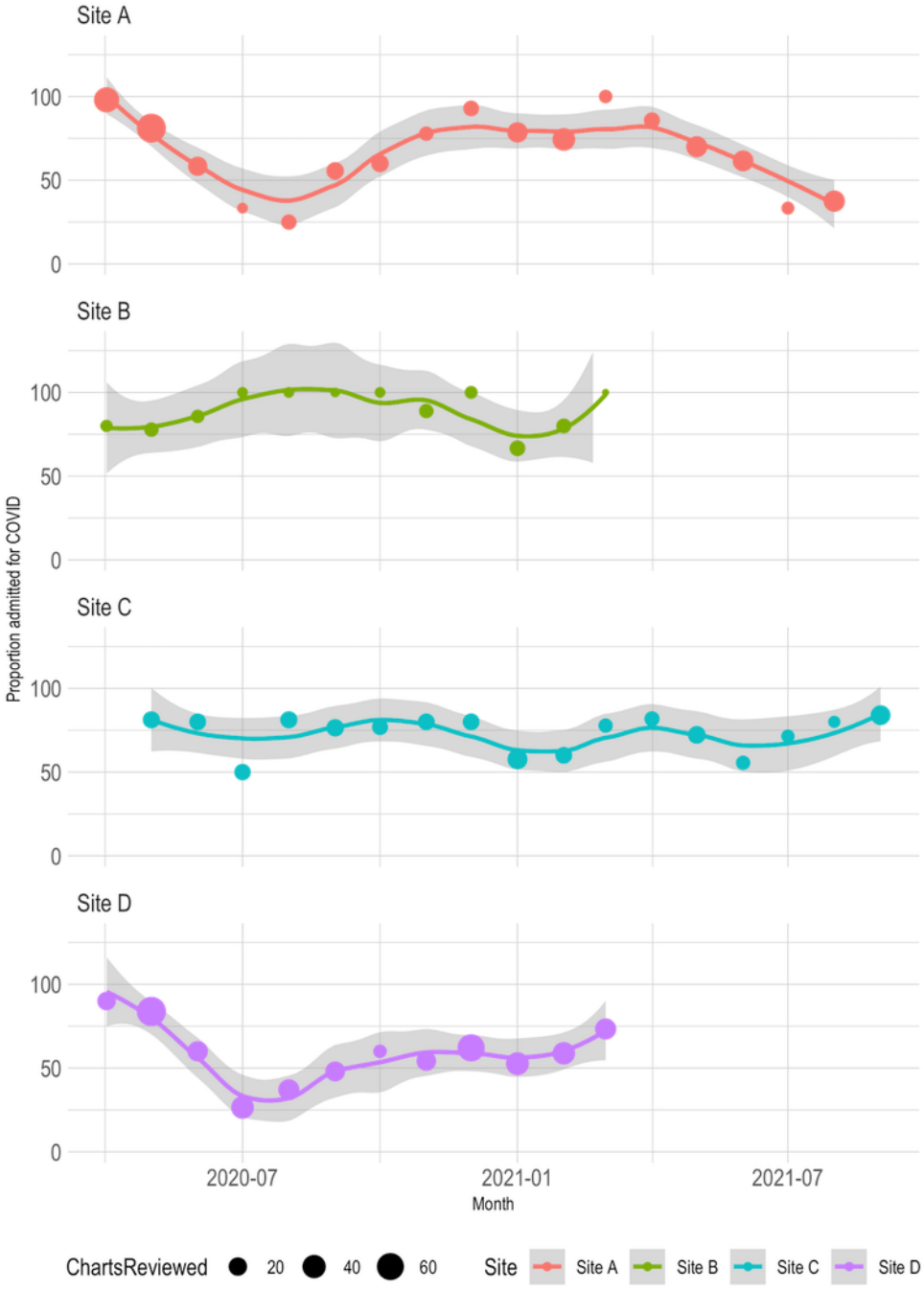
Table 4. Demographic characterization of the chart-reviewed cohort by site. For each row, the count and percentage (in parentheses) at each site are shown. Two sites did not report Hispanic/Latino. N values for each site are shown in the header; these might not exactly match the summation of each category due to blurring requirements.

Category	Site A (N=406), n (%)	Site B (N=70), n (%)	Site C (N=247), n (%)	Site D (N=400), n (%)
Age (years)				
0-25	14 (4)	11 (14)	4 (1)	11 (3)
26-49	95 (23)	15 (21)	26 (10)	76 (18)
50-69	138 (35)	22 (31)	99 (40)	135 (33)
70-79	72 (17)	9 (13)	59 (24)	90 (22)
80+	83 (20)	13 (18)	59 (24)	81 (19)
Race				
Asian	8 (2)	2 (3)	5 (2)	17 (4)
Black	60 (14)	9 (13)	58 (23)	97 (24)
Hispanic/Latino	21 (6)	N/A[a]	N/A	55 (14)
White	78 (19)	50 (71)	179 (72)	173 (42)
No information	230 (58)	8 (11)	5 (2)	61 (14)
Sex				
Male	200 (50)	42 (60)	121 (49)	188 (47)
Female	200 (50)	28 (40)	126 (51)	211 (52)

[a]N/A: not applicable.

XSL•FO
RenderX

Figure 3. Chart-reviewed proportion of admissions specifically for COVID-19 among all chart reviews by month at each site. The bubble size shows the relative number of patient chart reviews performed that month. The trendline was weighted by bubble size and was performed using locally weighted least squares (loess) regression. Note that the y axis and 95% CI limits extend above 100%.



XSL•FO
RenderX

Table 5. Top 10 ICD-10[a] diagnoses among patients' charts reviewed as admitted specifically for COVID-19, with the proportion of patients with each diagnosis at each site. Each patient might have multiple diagnoses, and therefore, the sum might be greater than 100%.

ICD-10 diagnosis	Site A (N=288), n (%)	Site B (N=59), n (%)	Site C (N=180), n (%)	Site D (N=240), n (%)
U07.1 Covid-19	265 (92)	54 (92)	145 (80)	226 (95)
J12.89 Other Viral Pneumonia	125 (44)	24 (41)	64 (35)	173 (70)
I10 Essential (Primary) Hypertension	113 (39)	16 (27)	74 (41)	89 (37)
J96.01 Acute Respiratory Failure With Hypoxia	75 (26)	20 (34)	56 (31)	139 (58)
E78.5 Hyperlipidemia, Unspecified	79 (28)	4 (7)	69 (38)	108 (46)
N17.9 Acute Kidney Failure, Unspecified	74 (25)	4 (7)	40 (22)	94 (39)
K21.9 Gastro-Esophageal Reflux Disease Without Esophagitis	64 (22)	<3 (<3)	57 (31)	65 (26)
Z87.891 Personal History of Nicotine Dependence	56 (18)	<3 (<3)	44 (24)	66 (27)
R09.02 Hypoxemia	81 (29)	15 (25)	21 (12)	43 (17)
J12.82 Pneumonia due to COVID-19	72 (25)	12 (20)	39 (22)	35 (15)

[a]ICD-10: International Classification of Diseases, Tenth Revision.

Table 6. Top 10 ICD-10[a] diagnoses among patients' charts reviewed as admitted with incidental COVID-19, with the proportion of patients with each diagnosis at each site. Each patient might have multiple diagnoses, and therefore, the sum might be greater than 100%.

ICD-10 diagnosis	Site A (N=85), n (%)	Site B (N=9), n (%)	Site C (N=54), n (%)	Site D (N=144), n (%)
U07.1 Covid-19	63 (74)	5 (56)	40 (73)	122 (85)
N17.9 Acute Kidney Failure, Unspecified	12 (14)	<3 (<11)	12 (22)	24 (17)
E11.22 Type 2 Diabetes Mellitus with Diabetic Chronic Kidney Disease	5 (6)	<3 (<11)	7 (13)	23 (15)
E11.9 Type 2 Diabetes Mellitus Without Complications	12 (11)	<3 (<11)	4 (7)	14 (11)
D64.9 Anemia, Unspecified	13 (19)	<3 (<11)	5 (9)	10 (6)
E87.2 Acidosis	8 (6)	<3 (<11)	<3 (<5)	12 (10)
J12.89 Other Viral Pneumonia	<3 (<2)	<3 (<11)	4 (7)	15 (12)
J96.01 Acute Respiratory Failure With Hypoxia	6 (8)	<3 (<11)	4 (7)	13 (8)
D69.6 Thrombocytopenia, Unspecified	5 (7)	<3 (<11)	6 (11)	12 (7)
N18.6 End-Stage Renal Disease	6 (7)	<3 (<11)	5 (9)	6 (5)

[a]ICD-10: *International Classification of Diseases, Tenth Revision*.

Phenotypes Using Hospital System Dynamics

Each site ran our HSD program to choose phenotypes of patients admitted *for COVID-19* versus patients admitted *incidentally with COVID-19*. The input of the program includes the chart-reviewed classifications and patient-level EHR data on the presence of 22 selected laboratory test types, 11 selected medication categories, 12 procedure categories, and all ICD-10 diagnosis codes that are dated within 48 hours of admission. This resulted in 1880 distinct features across all sites. (See Multimedia Appendix 1 for more information on the data dictionary.) The program selected 135 feature sets across all sites using these features. These were manually reduced to 32 (23.7%) by selecting the most predictive and removing duplicates and near-duplicates. These are summarized in Table 7, divided into phenotypes that use data that could be available immediately ("real time") and phenotypes using all data available after discharge ("retrospective"). We also reported the prevalence at each site among all SARS-CoV-2 PCR-positive hospitalizations (not just among chart-reviewed patients), which is the proportion of patients meeting the criteria of the feature sets.

XSL•FO
RenderX

Table 7. Top phenotyping feature sets by specificity, with a sensitivity of at least 0.60 for detecting admissions specifically for COVID-19. The table is grouped into feature sets involving potentially real-time data (laboratory tests) and all available data (presence of laboratory tests, medications, and diagnosis codes). Note that laboratory test results are not included in the feature sets. Ranges are shown in the summary statistics because multiple rules with similar performance were summarized using conjunctive normal form.

Phenotyping feature set	Site	Sensitivity	Specificity	Prevalence (%)
"Real-time" phenotypes (laboratory tests only)				
CRP[a] AND (Total Bilirubin OR Ferritin OR LDH[b]) AND (Lymphocyte Count OR Neutrophil Count) AND Cardiac Troponin	D	0.65-0.72	0.85	67-71
Ferritin AND LDH AND Cardiac Troponin AND (INR[c] OR PTT[d] OR Lymphocyte Count OR Neutrophil Count)	D	0.62-0.69	0.85	67-71
CRP AND (LDH AND/OR Ferritin) AND Cardiac Troponin	A	0.67-0.70	0.89-0.90	72-77
Procalcitonin OR D-dimer OR CRP OR Cardiac Troponin OR Ferritin	A	0.63-0.87	0.73-0.85	65-85
Any 2 of: Procalcitonin, LDH, CRP	B	0.56-0.58	0.67	63-67
D-dimer OR Ferritin OR CRP	C	0.26-0.37	0.86-0.93	54-58
"Retrospective" phenotypes (laboratory tests, medications, and diagnosis codes)				
Total bilirubin AND (Ferritin OR LDH OR Lymphocyte Count OR Neutrophil Count) AND diagnosis of Other Viral Pneumonia (J12.89)	D	0.62-0.64	0.92	46-48
Diagnosis of: Other Viral Pneumonia (J12.89) OR Acute Respiratory Failure with Hypoxia (J96.01) OR Anemia (D64.9)	D	0.70-0.74	0.82-0.88	50-63
Diagnosis of: Other Viral Pneumonia (J12.89) OR Supplemental Oxygen (severe)	D	0.75	0.82	61
CRP AND (LDH OR Ferritin) AND Cardiac Troponin	A	0.70	0.89	74-77
Remdesivir OR Procalcitonin OR Other Viral Pneumonia (J12.89) OR Nonspecific Abnormal Lung Finding (R91.8) OR Shortness of Breath (R06.02) OR Other COVID Disease (J12.82)	A	0.68-0.72	0.85-0.95	58-74
Hypoxemia (R09.02) OR Other Coronavirus as Cause of Disease (B97.29) OR Shortness of Breath (R06.02) OR Pneumonia (unspecified organism) (J18.9) OR Acute Respiratory Failure with Hypoxia (J96.01) OR Nonspecific Abnormal Lung Finding (R91.8)	B	0.63-0.68	0.89-0.99	54-67
D-dimer OR ferritin OR CRP OR Other Viral Pneumonia (J12.89) OR Acute Respiratory Failure with Hypoxia (J96.01)	C	0.71-0.75	0.79-0.86	52-58

[a]CRP: C-reactive protein.

[b]LDH: lactate dehydrogenase.

[c]INR: international normalized ratio.

[d]PTT: partial thromboplastin time.

We examined the top individual features over time at all sites. In the first half of 2020, a diagnosis of "Other Viral Pneumonia" (J12.89) was the only strong predictor of an admission specifically for COVID-19 across all 4 sites. In the second half of 2020, the phenotyping algorithm began selecting laboratory tests, including C-reactive protein (CRP), troponin, ferritin, and lactate dehydrogenase (LDH). In addition, the diagnosis "Other Coronavirus as Cause of Disease" (B97.29) began to be used at site B. By 2021, remdesivir and the diagnosis "Pneumonia due to COVID-19" (J12.82) additionally came into widespread use and became predictive of admissions specifically for COVID at site A.

Temporal Visualization of Phenotypes

We manually constructed 5 multisite phenotypes from elements in Table 7 that appeared at multiple sites. These were evaluated at each site: 2 variations of multisite diagnoses, 2 variations of all multisite features, and top laboratory tests. OR rules were favored due to better applicability across data sets (because of different coding practices at different sites), except for laboratory tests where the top pair of tests had high prevalence at every site. The best-performing phenotypes in each category are shown with their performance characteristics in Table 8, with the top single phenotype at each site in italics. In Figure 4, we plotted the performance of the top phenotype at each site (the boldfaced rows in Table 8) using the temporal phenotype visualization described in the Methods section. (The top phenotype involved all data types at every site except site C, where diagnoses alone performed better.)

XSL•FO
RenderX

Table 8. The best multisite phenotyping feature sets and their overall performance characteristics. The multisite phenotypes were derived from Table 7 by selecting components of phenotypes that appeared at multiple sites.

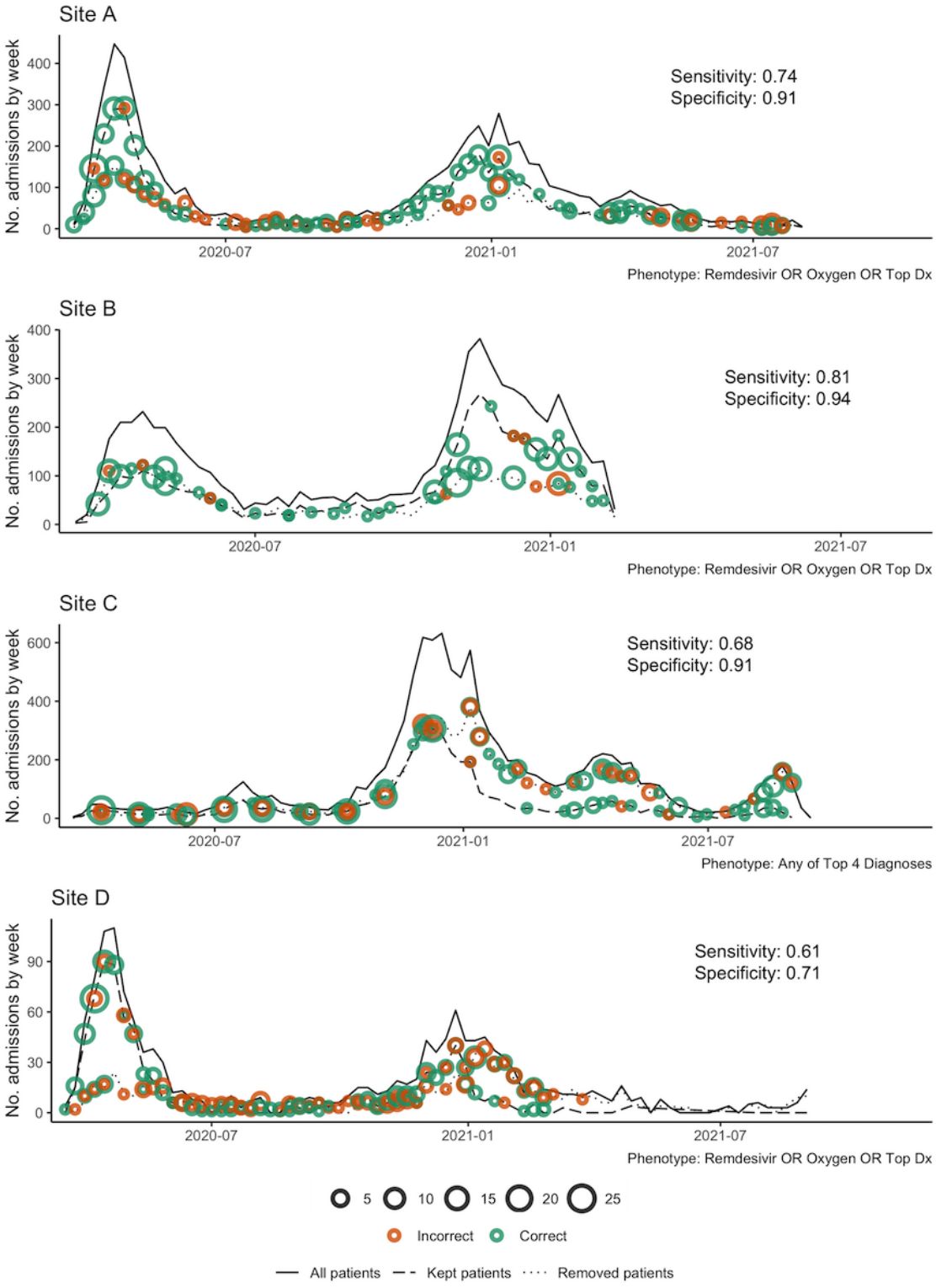
Phenotyping Feature Set	Description	Sensitivity, specificity
Other Viral Pneumonia OR Acute Respiratory Failure with Hypoxia OR Shortness of Breath OR Abnormal Lung Finding	*Retrospective phenotype:* Diagnoses mentioned in top feature sets at >1 site	• Site A: 0.79,0.72 • Site B: 0.88, 0.85 • *Site C: 0.69,0.90*[b] • Site D: 0.64,0.58
CRP[a] AND Ferritin	*Real-time phenotype:* Laboratory tests mentioned in top feature sets at all 4 sites	• Site A: 0.76,0.85 • Site B: 0.88, 0.85 • Site C: 0.42, 0.98 • Site D: 0.66, 0.55
Remdesivir OR Oxygen (severe) OR Dx of Other Viral Pneumonia	*Retrospective phenotype:* All items mentioned at multiple sites in OR feature sets	• *Site A: 0.74,0.91*[b] • *Site B: 0.81, 0.94*[b] • Site C: 0.60,0.92 • *Site D: 0.61,0.71*[b]

[a]CRP: C-reactive protein.

[b]The top-performing phenotype at each site is italicized.

XSL•FO
RenderX

Figure 4. Performance of the top phenotyping feature sets (Table 7) over time at each site. The y axis is the number of admissions per week, the x axis is the week, and overall sensitivity and specificity are shown on each figure panel. Solid lines show the total number of weekly admissions for patients with a positive SARS-CoV-2 PCR test. Dashed lines show the number of weekly admissions after filtering to select patients admitted specifically for COVID-19 (ie, removing all patients who do not meet the phenotyping feature set criteria). The dotted line shows the difference between the solid line and the dashed line (ie, patients removed from the cohort in the dashed line). Green dots indicate correct classification by the phenotype according to chart review. Orange dots indicate incorrect classification. The dot size is proportional to the number of chart reviews. PCR: polymerase chain reaction.



XSL•FO
RenderX

Discussion

Principal Results and Analysis

The COVID-19 pandemic has lasted for over 2 years, with multiple waves worldwide. Although hospital systems have been cyclically overwhelmed by patients seeking care for COVID-19, as health care systems began to open up before the second wave, elective surgeries were again performed starting in the later part of 2020, and especially in the second quarter of 2021, many approached the health care system for health issues (eg, accidents, strokes) while incidentally infected with SARS-CoV-2 [32]. This, along with the high false-positive rate of SARS-CoV-2 PCR tests in some situations [33-36], has led to increasing numbers of misclassified patients in analyses of COVID-19 characteristics and severity. This could be creating significant detection and reporting bias, leading to erroneous conclusions [10-13]. This study presents a multi-institutional characterization of 1123 hospitalized patients either incidentally infected with SARS-CoV-2 or specifically hospitalized for COVID-19 in 4 health care systems across multiple waves using consensus-based chart review criteria. Overall, we found that 764 (68%) of 1123 patients who tested SARS-CoV-2 positive were hospitalized because of COVID-19 but with significant variation during each wave of the pandemic.

We applied an item set–mining approach and established HSD principles to phenotype SARS-CoV-2 PCR-positive patients who were admitted specifically for COVID-19 by using data on charting patterns (eg, presence of laboratory tests within 48 hours of admission) rather than results (eg, laboratory results) [16,37]. HSD examines health care process data about a hospitalization, such as ordering/charting patterns, rather than the full data set. For example, to study severely ill patients, an HSD approach might select patients with a high total number of laboratory tests on the day of admission. This could be an indirect measure of clinical suspicion of disease complexity or severity. Previous work shows that proxies such as the total number of laboratory tests on the day of admission or the time of day of laboratory tests can be highly predictive of disease course [24,37]. Our methods sorted out who was treated for COVID-19 automatically, over time, with specificities above 0.70, even for some phenotypes discovered at a single site and applied to all 4. We focused on specificity because the goal was to remove false positives (ie, incidental SARS-CoV-2) from the cohort.

Our chart review protocol illustrates that patients who were admitted and had a positive SARS-CoV-2 PCR test were more likely to be admitted specifically for COVID-19 when disease prevalence was high (at least prior to Omicron). However, during periods in which health care systems were less restrictive (ie, resumed routine surgeries), a secondary measure/phenotype was critical for accurately classifying admissions specifically for SARS-CoV-2 infection.

As expected, we observed a lower proportion of hospitalizations specifically for COVID-19 in the summer months when disease prevalence was lower (Figure 3). One would expect this because there were fewer overall admissions as hospitals were recovering from the previous wave.

As expected, the top chart review criteria (Table 3) were *respiratory insufficiency* in admissions specifically for COVID-19 and *other* for incidental and uncertain admissions with SARS-CoV-2. Surprisingly, 10%-20% of patients admitted with incidental SARS-CoV-2 were diagnosed with pneumonia, respiratory failure, or acute kidney injury (Tables 5 and 6). This could reflect data collection issues, where some systems might repeat past problems automatically at hospital admission. In the case of codes for acute kidney injury, further investigation is needed to determine whether SARS-CoV-2-associated acute kidney injury (including COVID-19-associated nephropathy) occurs in patients we otherwise classified as having incidental admissions [38].

Health care systems are beginning to explore phenotyping feature sets to report admissions specifically for COVID-19. Starting January 2022 in Massachusetts, hospitals began reporting the number of for-COVID-19 hospitalizations as the count of admitted patients with both a SARS-CoV-2-positive test and a medication order for dexamethasone [22,23]. This simple phenotype was designed by the Massachusetts Department of Public Health as a first attempt, and it was based only on treatment recommendations for moderate-to-severe COVID-19 with hypoxia. It was not validated against a gold standard. Nonetheless, it illustrates the interest in EHR-based phenotyping for COVID-19.

Phenotypes with diagnosis codes tended to be the best-performing predictors of admissions, specifically for COVID-19. This could be because diagnosis codes represent either a clinically informed conclusion or a justification for ordering a test (implying the clinician suspected COVID-19). However, diagnoses are less prevalent in the population than laboratory tests and might not cover the entire population of admissions for COVID-19. Further, diagnoses early in hospitalization also do not always reflect the patient's eventual diagnosis or hospital-related complications that are more accurately reflected in discharge diagnoses. There was also some heterogeneity in the diagnoses used at different sites (eg, B97.29 "Other Coronavirus as Cause of Disease" was a top predictor only at site B). In addition, the presence of laboratory tests is useful for real-time detection systems because diagnosis codes usually are assigned after discharge. Clusters of tests for inflammatory markers (eg, LDH, CRP, and ferritin) appeared across most sites as predictive of hospitalizations, specifically for COVID-19, which fits intuitively because an underlying systemic pathophysiological mechanism of SARS-CoV-2 is thought to be an inflammatory process [39,40], and guidelines therefore have encouraged health care providers to check inflammatory markers on COVID-19 admissions [41,42]. Many of these inflammatory laboratory tests are not routinely ordered on all hospitalized patients and would therefore be expected to help distinguish COVID-19 from other diseases. However, laboratory protocol differences across sites may have reduced generalizability for this metric.

Our methods generated pairs of items using OR and groups of up to 4 using AND logical operators. Our feature sets were somewhat vulnerable to the problem that specificity decreases when multiple elements are combined with OR, although, in general, OR feature sets performed better across sites because

XSL•FO
RenderX

they could be designed to choose the top-performing elements at each site.

In addition to site differences, we also found changing disease management patterns over time. At the start of the pandemic, the only predictive phenotype was a pneumonia diagnosis. As standard COVID-19 order recommendations began to appear, laboratory orders became more consistent and predictive. Next, remdesivir began to be administered regularly. Finally, COVID-19-specific ICD-10 codes began to appear.

Overall, we found that an informatics-informed phenotyping approach successfully improved classification of for-COVID-19 versus incidental SARS-CoV-2-positive admissions, although generalizability was a challenge. Although some transfer learning is apparent (ie, a few phenotypes performed well across sites), local practice and charting patterns reduced generalizability. Specifically, phenotypes involving only laboratory tests did not perform well at site C, because the prevalence of these laboratory tests was low in the overall EHR data. This could be due to a data extraction or mapping issue in the underlying data warehouse. Site D had lower performance than other sites on the cross-site rules but not on the site-specific rules, perhaps highlighting less typical clinician treatment patterns.

Any of the multisite phenotypes developed here could be implemented as a cohort enhancement tool in hospital systems or data research networks, and the laboratory-only phenotypes ("CRP and Ferritin") could be used for real-time corrections in reporting. However, because of the changing nature of COVID-19 and practice and coding variation across sites, these phenotypes should be used primarily as a starting point. It is important to run the phenotyping algorithm on each individual site's data to tweak the rules to optimize them for each implementation.

Limitations

Although the current data start at the beginning of the pandemic, they do not include the current Omicron wave nor much of the Delta wave. We believe that the techniques introduced here (if not the phenotypes themselves) will be applicable to these variants, and we are planning future studies to validate this.

Our phenotypes demonstrated some transfer learning but not enough to create a single phenotype applicable to all sites. Technically, our system used machine learning at individual sites, but results were manually aggregated across sites. Emerging techniques for federated learning [43] might reduce the manual work required and increase the complexity of possible cross-site phenotype testing.

Finally, an inherent weakness of EHR-based research is that EHR data do not directly represent the state of the patient, because some observations are not recorded in structured data and some entries in the EHR are made for nonclinical reasons (eg, to justify the cost of a test or to ensure adequate reimbursement for services). This is common to all EHR research efforts, and we mitigated this limitation by developing chart-verified phenotypes.

Conclusion

At 4 health care systems around the United States over an 18-month period starting in March 2020, we developed and applied standardized chart review criteria to characterize the correct classification of hospitalization specifically for COVID-19 as compared to incidental hospitalization of a patient with a positive SARS-CoV-2 test or ICD-10 code. Then we applied HSD and frequent item set mining to electronic phenotyping to generate phenotypes specific to hospitalizations for COVID-19, and we showed how patterns changed over the course of the pandemic. Application of this approach could improve public health reporting, health care system resource disbursement, and research conclusions.

Acknowledgments

We would like to additionally thank Trang Le, PhD, for her assistance in developing the temporal phenotype visualization R script; Karen Olson, PhD, for her knowledge of the R programming language and insights into the topic of this study; Margaret Vella for her administrative coordination; and Isaac Kohane, MD, PhD, for convening, leading, and guiding the Consortium for Clinical Characterization of COVID-19 by electronic health record (4CE).

We would like to express our gratitude to all group members of the Consortium for Clinical Characterization of COVID-19 by EHR (4CE) for their contribution to the study [Multimedia Appendix 2].

JSM is supported by the National Institutes of Health (NIH)/National Library of Medicine (NLM) T15LM007092. MM is supported by NIH/National Center for Advancing Translational Sciences (NCATS) UL1TR001857. AMS is supported by NIH/National Heart, Lung, and Blood Institute (NHLBI) K23HL148394 and L40HL148910 and NIH/NCATS UL1TR001420. GMW is supported by NIH/NCATS UL1TR002541, NIH/NCATS UL1TR000005, NIH/NLM R01LM013345, and NIH/National Human Genome Research Institute (NHGRI) 3U01HG008685-05S2. WY is supported by NIH T32HD040128. KBW is supported by NIH/NHLBI R01 HL151643-01. YL is supported by NIH/NCATS U01TR003528 and NLM 1R01LM013337. GSO is supported by NIH U24CA210867 and P30ES017885. SV is supported by NCATS UL1TR001857. JHH is supported by NCATS UL1-TR001878. ZX is supported by the National Institute of Neurological Disorders and Stroke (NINDS) R01NS098023 and NINDS R01NS124882. SNM is supported by NCATS 5UL1TR001857-05 and NHGRI 5R01HG009174-04.

The study sponsors played no role in defining or designing the study, nor did they play any role in collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the manuscript for publication.

XSL•FO
RenderX

Data Availability

All data analysis code developed for this study is available under the Mozilla Public License v2 with a health care disclaimer.

The electronic health record (EHR) data sets analyzed during this study cannot be made publicly available due to regulations for protecting patient privacy and confidentiality. These regulations also prevent the data from being made available upon request from the authors. Any questions about the data set can be directed to the corresponding author.

Authors' Contributions

JGK and SNM conceptualized and designed the study. ZHS, GAB, JGK, and SNM conceptualized and designed the chart review process. JGK, ZHS, MRH, CJK, JSM, MM, MJS, ACP, GMW, WY, YL, SV, ZX, and GAB contributed to data collection (of electronic health record [EHR] data or chart reviews). JGK, ZHS, MRH, CJK, JSM, MM, MJS, ACP, AMS, GMW, WY, PA, KBW, GSO, SV, JHH, ZX, GAB, and SNM contributed to data analysis or interpretation. JGK, ZHS, MRH, CJK, MM, HE, AMS, GMW, PA, KBW, YL, GSO, JHH, ZX, GAB, and SNM contributed to drafting and revision of the manuscript. SNM contributed to grant funding. All authors approved the final draft of the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest. JGK reports a consulting relationship with the i2b2-tranSMART Foundation through Invocate, Inc. CJK reports consulting for the University of California, Berkeley; the University of Southern California (USC), and the University of California, San Francisco (UCSF). AMS reports funding from the National Institutes of Health (NIH)/National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) R01DK127208, NIH/National Heart, Lung, and Blood Institute (NHLBI) R01HL146818, and institutional pilot awards from the Wake Forest School of Medicine. GMW reports consulting for the i2b2-tranSMART Foundation. PA reports consulting for the Cincinnati Children's Hospital Medical Center (CCHMC) and Boston Children's Hospital (BCH). ZX has received research support from the NIH, the Department of Defense, and Octave Biosciences and has served on the scientific advisory board for Genentech/Roche. SNM reports professional relationships with the Scientific Advisory Board for Boston University, the Universidad de Puerto Rico, the University of California, Los Angeles (UCLA), the University of Massachusetts Medical School (UMMS), and the Kenner Family Research Fund.

Multimedia Appendix 1
4CE data dictionary. 4CE: Consortium for Clinical Characterization of COVID-19 by EHR. EHR: electronic health record.
[PDF File (Adobe PDF File), 278 KB - jmir_v24i5e37931_app1.pdf]

Multimedia Appendix 2
List of group members of the Consortium for Clinical Characterization of COVID-19 by EHR (4CE).
[DOCX File , 18 KB - jmir_v24i5e37931_app2.docx]

References

1. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
2. Brat GA, Weber GM, Gehlenborg N, Avillach P, Palmer NP, Chiovato L, et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. NPJ Digit Med 2020;3:109 [FREE Full text] [doi: 10.1038/s41746-020-00308-0] [Medline: 32864472]
3. 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;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
4. Visweswaran S, Samayamuthu MJ, Morris M, Weber GM, MacFadden D, Trevvett P, et al. Development of a coronavirus disease 2019 (COVID-19) application ontology for the Accrual to Clinical Trials (ACT) network. JAMIA Open 2021 Apr;4(2):ooab036 [FREE Full text] [doi: 10.1093/jamiaopen/ooab036] [Medline: 34113801]
5. Khullar D. Do the Omicron Numbers Mean What We Think They Mean? The New Yorker. 2022 Jan 16. URL: https://www.newyorker.com/magazine/2022/01/24/do-the-omicron-numbers-mean-what-we-think-they-mean [accessed 2022-01-25]
6. Centers for Disease Control and Prevention. ICD-10-CM Official Coding and Reporting Guidelines. National Center for Health Statistics. 2020. URL: https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf [accessed 2022-04-29]
7. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, National COVID Cohort Collaborative (N3C) Consortium. Clinical characterization and prediction of clinical severity of SARS-CoV-2 infection among US adults using data from the US National COVID Cohort Collaborative. JAMA Netw Open 2021 Jul 01;4(7):e2116901 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.16901] [Medline: 34255046]

XSL•FO
RenderX

8. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. Nature 2021 Jun;594(7862):259-264. [doi: 10.1038/s41586-021-03553-9] [Medline: 33887749]
9. National Notifiable Diseases Surveillance System (NNDSS), Centers for Disease Control and Prevention. Coronavirus Disease 2019 (COVID-19) 2021 Case Definition. 2021. URL: https://ndc.services.cdc.gov/case-definitions/coronavirus-disease-2019-2021/ [accessed 2022-04-29]
10. Sah P, Fitzpatrick MC, Zimmer CF, Abdollahi E, Juden-Kelly L, Moghadas SM, et al. Asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis. Proc Natl Acad Sci U S A 2021 Aug 24;118(34):e2109229118 [FREE Full text] [doi: 10.1073/pnas.2109229118] [Medline: 34376550]
11. Buitrago-Garcia D, Egli-Gany D, Counotte MJ, Hossmann S, Imeri H, Ipekci AM, et al. Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: a living systematic review and meta-analysis. PLoS Med 2020 Sep;17(9):e1003346 [FREE Full text] [doi: 10.1371/journal.pmed.1003346] [Medline: 32960881]
12. Fisman DN, Tuite AR. Asymptomatic infection is the pandemic's dark matter. Proc Natl Acad Sci U S A 2021 Sep 21;118(38):e2114054118 [FREE Full text] [doi: 10.1073/pnas.2114054118] [Medline: 34526404]
13. Cohen JB, D'Agostino McGowan L, Jensen ET, Rigdon J, South AM. Evaluating sources of bias in observational studies of angiotensin-converting enzyme inhibitor/angiotensin II receptor blocker use during COVID-19: beyond confounding. J Hypertens 2021 Apr 01;39(4):795-805 [FREE Full text] [doi: 10.1097/HJH.0000000000002706] [Medline: 33186321]
14. Garrett N, Tapley A, Andriesen J, Seocharan I, Fisher LH, Bunts L, et al. High rate of asymptomatic carriage associated with variant strain omicron. medRxiv 2022 Jan 14 [FREE Full text] [doi: 10.1101/2021.12.20.21268130] [Medline: 35043118]
15. Yu S, Ma Y, Gronsbell J, Cai T, Ananthakrishnan AN, Gainer VS, et al. Enabling phenotypic big data with PheNorm. J Am Med Inform Assoc 2018 Jan 01;25(1):54-60 [FREE Full text] [doi: 10.1093/jamia/ocx111] [Medline: 29126253]
16. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. J Am Med Inform Assoc 2013 Jan 01;20(1):117-121 [FREE Full text] [doi: 10.1136/amiajnl-2012-001145] [Medline: 22955496]
17. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer V, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
18. Klann JG, Estiri H, Weber GM, Moal B, Avillach P, Hong C, Consortium for Clinical Characterization of COVID-19 by EHR (4CE) (CONSORTIA AUTHOR), et al. Validation of an internationally derived patient severity phenotype to support COVID-19 analytics from electronic health record data. J Am Med Inform Assoc 2021 Jul 14;28(7):1411-1420 [FREE Full text] [doi: 10.1093/jamia/ocab018] [Medline: 33566082]
19. 4CE: Consortium for Clinical Characterization of COVID-19 by EHR. URL: https://covidclinical.net/ [accessed 2022-04-29]
20. Weber GM, Zhang HG, L'Yi S, Bonzel C, Hong C, Avillach P, et al. International changes in COVID-19 clinical trajectories across 315 hospitals and 6 countries: retrospective cohort study. J Med Internet Res 2021 Oct 11;23(10):e31400 [FREE Full text] [doi: 10.2196/31400] [Medline: 34533459]
21. Kohane IS, Aronow BJ, Avillach P, Beaulieu-Jones BK, Bellazzi R, Bradford RL, Consortium For Clinical Characterization Of COVID-19 By EHR (4CE), et al. What every reader should know about studies using electronic health record data but may be afraid to ask. J Med Internet Res 2021 Mar 02;23(3):e22219 [FREE Full text] [doi: 10.2196/22219] [Medline: 33600347]
22. Fatima S. Mass. Hospitals Will Begin Reporting Primary vs. Incidental COVID-19 Admissions on Monday, DPH Says. The Boston Globe. 2022. URL: https://www.bostonglobe.com/2022/01/06/nation/teachers-union-accuses-state-gross-incompetence-questions-swirl-around-masks-distributed-school-staff/ [accessed 2022-01-26]
23. Bebinger M. State Changes COVID Reporting to Distinguish between Primary and Incidental Hospital Cases. WBUR. 2022. URL: https://www.wbur.org/news/2022/01/21/massachusetts-primary-incidental-coronavirus-grouping [accessed 2022-01-26]
24. Agniel D, Kohane IS, Weber GM. Biases in electronic health record data due to processes within the healthcare system: retrospective observational study. BMJ 2018 Apr 30;361:k1479 [FREE Full text] [doi: 10.1136/bmj.k1479] [Medline: 29712648]
25. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. 1993 Jun Presented at: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data; 1993; New York, NY p. 207-216. [doi: 10.1145/170036.170072]
26. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules.: IBM Almaden Research Center URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.40.7506&rep=rep1&type=pdf [accessed 2022-04-29]
27. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. J Biomed Inform 2010 Dec;43(6):891-901 [FREE Full text] [doi: 10.1016/j.jbi.2010.09.009] [Medline: 20884377]
28. Klann JG, Szolovits P, Downs SM, Schadow G. Decision support from local data: creating adaptive order menus from past clinician behavior. J Biomed Inform 2014 Apr;48:84-93 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.005] [Medline: 24355978]

XSL•FO
RenderX

29. Adamo J. Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms. Berlin/Heidelberg, Germany: Springer Science & Business Media; 2001.
30. Klann J. Jeff Klann's 4CE-Related Tools Github. URL: https://github.com/jklann/jgk-i2b2tools [accessed 2022-03-16]
31. Klann JG, Joss M, Shirali R, Natter M, Schneeweiss S, Mandl KD, et al. The ad-hoc uncertainty principle of patient privacy. AMIA Jt Summits Transl Sci Proc 2018;2017:132-138 [FREE Full text] [Medline: 29888058]
32. CMS Releases Recommendations on Adult Elective Surgeries, Non-Essential Medical, Surgical, and Dental Procedures During COVID-19 Response. CMS Newsroom. 2020. URL: https://www.cms.gov/newsroom/press-releases/cms-releases-recommendations-adult-elective-surgeries-non-essential-medical-surgical-and-dental [accessed 2022-04-29]
33. Lan L, Xu D, Ye G, Xia C, Wang S, Li Y, et al. Positive RT-PCR test results in patients recovered from COVID-19. JAMA 2020 Apr 21;323(15):1502-1503 [FREE Full text] [doi: 10.1001/jama.2020.2783] [Medline: 32105304]
34. Wu J, Liu X, Liu J, Liao H, Long S, Zhou N, et al. Coronavirus disease 2019 test results after clinical recovery and hospital discharge among patients in China. JAMA Netw Open 2020 May 01;3(5):e209759 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.9759] [Medline: 32442288]
35. Skittrall JP, Wilson M, Smielewska AA, Parmar S, Fortune MD, Sparkes D, et al. Specificity and positive predictive value of SARS-CoV-2 nucleic acid amplification testing in a low-prevalence setting. Clin Microbiol Infect 2021 Mar;27(3):469.e9-469.e15 [FREE Full text] [doi: 10.1016/j.cmi.2020.10.003] [Medline: 33068757]
36. Mina MJ, Peto TE, García-Fiñana M, Semple MG, Buchan IE. Clarifying the evidence on SARS-CoV-2 antigen rapid tests in public health responses to COVID-19. Lancet 2021 Apr 17;397(10283):1425-1427 [FREE Full text] [doi: 10.1016/S0140-6736(21)00425-6] [Medline: 33609444]
37. Weber GM, Kohane IS. Extracting physician group intelligence from electronic health records to support evidence based medicine. PLoS One 2013;8(5):e64933 [FREE Full text] [doi: 10.1371/journal.pone.0064933] [Medline: 23734227]
38. Batlle D, Soler MJ, Sparks MA, Hiremath S, South AM, Welling PA, COVID-19ACE2 in Cardiovascular, Lung, Kidney Working Group. Acute kidney injury in COVID-19: emerging evidence of a distinct pathophysiology. J Am Soc Nephrol 2020 Jul;31(7):1380-1383 [FREE Full text] [doi: 10.1681/ASN.2020040419] [Medline: 32366514]
39. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
40. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, China Medical Treatment Expert Group for Covid-19. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032] [Medline: 32109013]
41. Kim A, Gandhi RT. COVID-19: Management in Hospitalized Adults. UpToDate. 2022. URL: https://www.uptodate.com/contents/covid-19-management-in-hospitalized-adults?search=covid-19%20admission%20management&source=covid19_landing&usage_type=main_section [accessed 2022-04-29]
42. Centers for Disease Control and Prevention. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html [accessed 2022-02-04]
43. Pfitzner B, Steckhan N, Arnrich B. Federated learning in a medical context: a systematic literature review. ACM Trans Internet Technol 2021 Jun;21(2):1-31. [doi: 10.1145/3412357]

Abbreviations

4CE: Consortium for Clinical Characterization of COVID-19 by EHR
BIDMC: Beth Israel Deaconess Medical Center
CRP: C-reactive protein
EHR: electronic health record
HSD: hospital system dynamics
ICD-10: International Classification of Diseases, Tenth Revision
IRB: Institutional Review Board
INR: international normalized ratio
LDH: lactate dehydrogenase
MGB: Mass General Brigham
NWU: Northwestern University
PCR: polymerase chain reaction
PPV: positive predictive value
UPITT: University of Pittsburgh/University of Pittsburgh Medical Center

XSL•FO
RenderX

Edited by C Basch; submitted 16.03.22; peer-reviewed by P Stephan, A King; comments to author 11.04.22; revised version received 22.04.22; accepted 22.04.22; published 18.05.22.

Please cite as:
Klann JG, Strasser ZH, Hutch MR, Kennedy CJ, Marwaha JS, Morris M, Samayamuthu MJ, Pfaff AC, Estiri H, South AM, Weber GM, Yuan W, Avillach P, Wagholikar KB, Luo Y, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Omenn GS, Visweswaran S, Holmes JH, Xia Z, Brat GA, Murphy SN
Distinguishing Admissions Specifically for COVID-19 From Incidental SARS-CoV-2 Admissions: National Retrospective Electronic Health Record Study
J Med Internet Res 2022;24(5):e37931
URL: https://www.jmir.org/2022/5/e37931
doi:10.2196/37931
PMID:35476727

©Jeffrey G Klann, Zachary H Strasser, Meghan R Hutch, Chris J Kennedy, Jayson S Marwaha, Michele Morris, Malarkodi Jebathilagam Samayamuthu, Ashley C Pfaff, Hossein Estiri, Andrew M South, Griffin M Weber, William Yuan, Paul Avillach, Kavishwar B Wagholikar, Yuan Luo, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), Gilbert S Omenn, Shyam Visweswaran, John H Holmes, Zongqi Xia, Gabriel A Brat, Shawn N Murphy. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 18.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Effect of Dissemination Pathways on Uptake and Relative Costs for a Transdiagnostic, Self-guided Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation: Comparative Implementation Study

Philip J Batterham[1], BSc Psychology (Hons), MPH, PhD; Amelia Gulliver[1], BA, BScPsych (Hons), PhD; Ella Kurz[1], BMid, PhD; Louise M Farrer[1], BPsych (Hons), PhD; Christiaan Vis[2], MA; Josien Schuurmans[3], PhD; Alison L Calear[1], BAppPsych (Hons), PhD

[1]Centre for Mental Health Research, The Australian National University, Canberra, Australia

[2]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[3]GGZ InGeest, Amsterdam, Netherlands

Corresponding Author:
Philip J Batterham, BSc Psychology (Hons), MPH, PhD
Centre for Mental Health Research
The Australian National University
63 Eggleston Rd
Canberra, 2601
Australia
Phone: 61 261251031
Email: philip.batterham@anu.edu.au

Abstract

Background: Self-guided web-based programs are effective; however, inadequate implementation of these programs limits their potential to provide effective and low-cost treatment for common mental health problems at scale. There is a lack of research examining optimal methods for the dissemination of web-based programs in the community.

Objective: This study aimed to compare the uptake, reach, relative costs, and adherence associated with 3 community-based pathways for delivering a low-intensity web-based transdiagnostic mental health program. The 3 dissemination pathways were social media advertising, advertising in general practice, and advertising in pharmacies.

Methods: Participants were recruited on the web, from general practices, or from community pharmacies; completed a screener for psychological distress; and were offered the 4-week *FitMindKit* program—a 12-module psychotherapeutic intervention. Uptake was defined as the number of participants who enrolled in the web-based program; reach was defined as the rate of uptake per exposure; and costs were calculated based on staff time, equipment, and advertising. Adherence was assessed as the number of modules of *FitMindKit* completed by the participants.

Results: Uptake comprised 1014 participants who were recruited through the 3 dissemination pathways: on the web (991/1014, 97.73%), in general practice (16/1014, 1.58%), and in pharmacy (7/1014, 0.69%). Reach was highest for social media: 1 in every 50 people exposed to web-based advertising took up the intervention compared with 1 in every 441 in general practitioner clinics and 1 in every 1708 in pharmacies. The dissemination cost was US $4.87 per user on social media, US $557 per user for general practitioner clinics, and US $1272 per user for pharmacy dissemination. No significant differences in adherence were observed between the conditions, whereas all pathways showed an underrepresentation of men and linguistic diversity.

Conclusions: The web-based dissemination pathway was the most efficient and cost-effective for delivering a self-guided internet-based mental health program to people in the community. More research is needed to identify how best to engage men and those with culturally diverse backgrounds in web-based interventions.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618001688279; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=376113

(J Med Internet Res 2022;24(5):e34769) doi:10.2196/34769

XSL•FO
RenderX

KEYWORDS

implementation; mental health; adherence; uptake; internet

Introduction

Background

Depression and anxiety are common mental health problems, with 10% to 15% of the global population experiencing a lifetime mood disorder and 10% to 15% experiencing a lifetime anxiety disorder, and depression and anxiety prevalence worldwide were 3152 and 4802 per 100,000, respectively, in 2020 [1-4]. They can cause significant disability burden and affect many domains of life, including physical, social, work, and education [5-7]. Despite this considerable impact, more than half of people experiencing mental disorders worldwide do not seek professional help at onset, typically delaying care for many years, with <10% receiving adequate treatment [8-10]. A number of factors have been identified as contributing to the low levels of help seeking in the community, including the accessibility of services (location and cost), mental health stigma, and mental health literacy [11-13]. Web-based self-guided mental health interventions have been proposed as a way of addressing low help-seeking rates, particularly as they can circumvent some of the barriers by providing private, low-cost access to treatment [14]. Consequently, community-based provision of evidence-based, low-intensity, self-guided web-based mental health programs may assist in increasing treatment access for people who would not otherwise seek help [15,16].

In addition to being prevalent, depression and anxiety also have high rates of comorbidity between them and with other mental disorders [17]. Transdiagnostic therapeutic approaches involve treating multiple conditions simultaneously by targeting the core mechanisms underlying comorbid conditions (eg, negative thinking patterns) rather than diagnosis-specific treatment [18]. Transdiagnostic programs are highly suited for web-based dissemination and have been demonstrated to be effective in reducing symptoms of depression and anxiety [19,20]. In addition, self-guided interventions that do not require clinician support have been found to be safe and effective in reducing the symptoms of depression and anxiety [15,21,22], particularly for people with mild to moderate symptom severity. Self-guided interventions require minimal resources and are consequently highly scalable [23], although clinician-guided programs typically show stronger effects [15] and greater acceptability [24]. Despite the availability and accessibility of effective self-guided interventions, community uptake of web-based mental health programs remains poor [14,25]. For example, a naturalistic study examining uptake in the community of a web-based depression program found that half of all visitors to the program's website (N=194,840) did not register for the program [26]. Therefore, ensuring that effective programs are implemented in the community is critical to realizing the full potential of web-based mental health programs.

The field of implementation science focuses on how best to implement evidence-based treatments into health care services [27]. In addition to low uptake within the community, the implementation of web-based mental health programs in clinical service settings remains a challenge [16,28]. Although there is strong evidence that both self-guided and clinician-guided programs are effective for preventing and treating depression and anxiety [16,28-31], only 14% of new evidence-based interventions enter routine practice [32,33]. In addition, successful programs are typically only implemented several years after establishing their clinical efficacy [32,34].

The uptake of web-based mental health programs within primary care (eg, via general practitioners [GPs]) is emerging but has been relatively slow [35-37]. Previous research conducted in Australia [38] examined the potential of recruiting people with type 2 diabetes and depression into a web-based mental health program. This study found that recruitment through general practice (n=24 participants) was unsuccessful compared with going through registries and word of mouth (n=196) or web-based recruitment methods (eg, Facebook; n=520) [38]. The authors identified staff attitudes toward internet interventions as a key barrier to recruitment through primary care in Australia [38]. Other potential community-based settings that offer promise in increasing uptake of internet interventions include community pharmacies (ie, pharmacies outside of hospitals and other care facilities), which have previously been investigated as a potential avenue for recruitment of participants into community health care trials [39]. However, little is known about the potential of pharmacies to act as a pathway to treatment for web-based mental health services. Dissemination of internet interventions directly to end users through web-based methods, such as social media marketing or digital apothecaries (directories of evidence-based programs), is more common and may be better suited to the modality and self-directed nature of these interventions [40].

The gap between the evidence and uptake of web-based mental health programs is exacerbated by a paucity of translational research into how best to implement these programs within traditional health care settings and directly in the community. Conducted as part of a broader project targeting implementation [41], this project addresses this research gap by testing different pathways for disseminating the *FitMindKit* e-mental health program to people experiencing symptoms of depression and anxiety in the community. *FitMindKit* is a self-guided, web-based mental health program that delivers low-intensity transdiagnostic cognitive behavioral therapy over 4 weeks via 12 modules comprising brief videos and self-directed exercises [42,43]. The program has been found to be effective in a community-based trial for improving symptoms of depression, panic disorder, and social anxiety [44]. In this study, dissemination was tested as a specific implementation strategy [45] to increase intervention reach and engage more people in need of treatment with a low-threshold web-based intervention. We compared the 3 dissemination pathways to identify which had the greatest reach, uptake, and adherence and lowest costs.

XSL•FO
RenderX

Hypotheses

We hypothesized that direct social media advertising on the web would be associated with greater uptake (number of participants screened and enrolled in the web-based program) than recruitment via advertising using posters or tablet computers in general practices and pharmacies. It was anticipated that the relative reach (ratio of uptake to exposure) would be highest in the social media pathway, whereas the cost per enrolled participant would consequently be lower in this pathway. In addition, we hypothesized that engagement with the program (based on adherence; module completion) would be lowest among people recruited through social media, as this group had the least possibility of human contact in the recruitment process [46]; that is, the higher uptake on the web may come at the expense of engagement with the intervention. We did not have any specific hypotheses regarding outcomes for the pharmacy pathway, as this pathway has not been investigated previously in e-mental health implementation research. We also examined the diversity of participants recruited via each dissemination arm to assess the relative inclusivity of each dissemination pathway as an exploratory analysis with no specific hypotheses.

Methods

Study Design and Participants

This project was part of a larger implementation trial [41] and was designed to test the effectiveness of three different dissemination pathways—(1) social media, (2) general practices, and (3) pharmacies—in delivering an e-mental health program (*FitMindKit*) to people in the community with elevated symptoms of psychological distress. Dissemination methods were tailored for each pathway, with different approaches to inviting participants to complete a brief screening measure assessing psychological distress. The tailoring involved consulting with clinic and pharmacy staff about the most appropriate ways of engaging their clients and optimizing social media advertising based on past evidence [47].

Adults aged ≥18 years who were screened with elevated psychological distress based on scores of 8 to 17 on the Distress Questionnaire 5 (DQ5) [48], which primarily reflects symptoms of mood or anxiety disorders [49], were eligible and invited to use *FitMindKit* for a period of 4 weeks. The DQ5 is a brief screening measure comprising 5 items describing symptoms of common mental disorders. Participants endorsed the frequency of each item over the past 30 days on a 5-point scale ranging from never (1) to always (5). Total scale scores ranged from 5 to 25; low scores (5-7) indicate no or low psychological distress, scores of 8 to 17 indicate a moderate level of psychological distress, and scores of 18 to 25 indicate a high level of psychological distress, based on quartiles from a previous validation study [48]. Those who were ineligible because they were in the low-risk category (DQ5 score=5-7) were provided with feedback to continue monitoring their psychological well-being and with help resources to access if their symptoms changed. Those who were ineligible because they were in the high-risk category (DQ5 score=18-25) were provided with feedback strongly encouraging them to seek help from a health professional and contact details for face-to-face, telephone-based, and web-based mental health resources and services.

Participants completed brief pre- and postprogram questionnaires and received weekly email reminders to use *FitMindKit*. The study was conducted from July 2018 to December 2020; however, the social media arm was only active from October 2018 to August 2019; therefore, only data from this period were used for comparisons for all pathways.

Ethics Approval

The ethical aspects of this research were approved by the Australian National University Human Research Ethics Committee (protocol number 2017/911).

Recruitment

General Practice and Pharmacy

General practices and pharmacies located in the Australian Capital Territory were invited to participate in the study via an invitation letter, followed up by a telephone call and an in-person visit. Invitations were sent to 49 general practices and 41 pharmacies. Of these, 10% (5/49) of general practices and 12% (5/41) of pharmacies participated in the study, with the remainder either not responding after 2 follow-ups or declining because of lack of interest, lack of time, or limited infrastructure (eg, no Wi-Fi connection). Promotional posters and flyers were provided to participating general practices and pharmacies to promote the study to patients and customers. A tablet computer was also provided in the waiting areas of general practices and pharmacies, where interested participants could read information about the study, provide their consent to participate, and complete a screening survey to assess their eligibility for the study.

During the project, the researchers actively engaged with staff in the participating general practices and pharmacies to identify and remediate any barriers or difficulties they faced in implementing the project. Key staff in practices and pharmacies were also asked to complete a brief measure at the beginning and end of the project to assess their experience with the implementation process.

Social Media

Participants were recruited to the social media pathway of the study via Facebook and Instagram advertisements. To ensure that the geographical catchment area for this pathway was consistent with that of the GP clinics and pharmacies, the advertisement specifically targeted adults living in the Australian Capital Territory aged ≥18 years. It read, “Want to learn more about your mental health and well-being? Complete a brief survey and 4 week online program. Participants needed for a mental health study.” A series of images that were used on a rotational basis accompanied the advertisements and featured natural scenes (eg, trees, plants, and clouds). Participants who clicked on the advertisement received a full information sheet with details of the program, were screened, completed baseline measures, and then randomly allocated to receive *FitMindKit* (991/1986, 49.9%) or an attention control website—*HealthWatch* (995/1986, 50.1%)—for 4 weeks. This

XSL•FO
RenderX

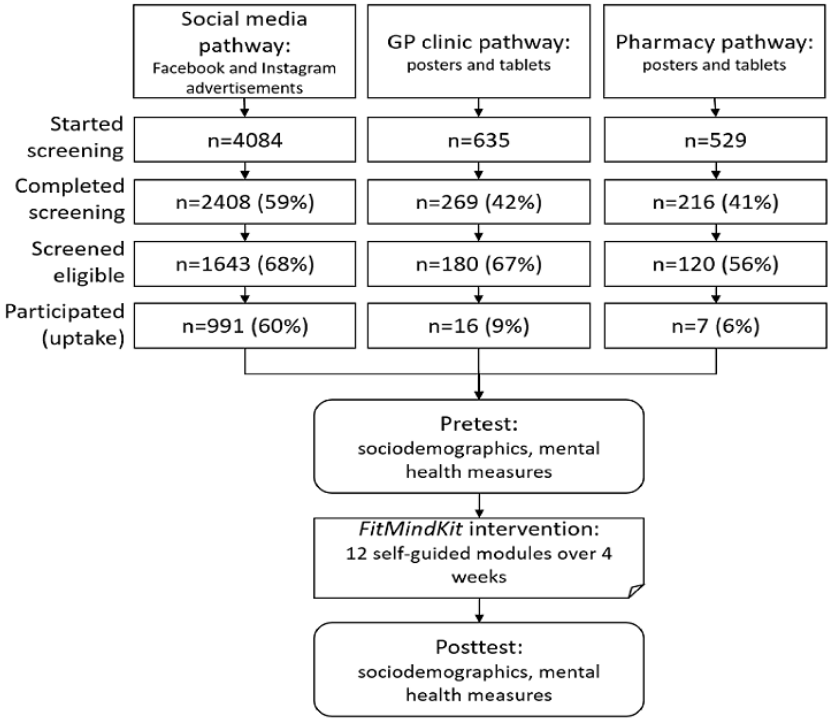
study only included data from participants who were allocated to the *FitMindKit* condition and excluded participants who were allocated to the control condition.

Procedure

Overview

The flow of participants through the study is summarized in Figure 1. Data for the current analyses were obtained from screening and pretest assessments only. After obtaining informed consent and screening, participants were directed to create an account in the project portal using an email address and password. Within the portal, all participants completed a baseline questionnaire, irrespective of the dissemination pathway. Those in the GP and pharmacy pathways were offered the intervention, and those in the social media pathway were randomized to receive either the intervention or an attention control condition, although only data from participants who received the active intervention were included in the current analyses. Participants were sent a weekly reminder email to encourage them to engage with *FitMindKit*. Following the 4-week period, participants were sent an email inviting them to complete the postintervention assessment. They received 2 reminder emails if they did not complete the postintervention assessment after 1 and 2 weeks.

Figure 1. Flow of participants in the study across the 3 dissemination pathways. GP: general practitioner.



FitMindKit Program

FitMindKit is a transdiagnostic web-based program comprising brief (10-minute) therapeutic modules to reduce the symptoms of common mental disorders. Each module comprises a brief video (2-6 minutes) using a series of professionally designed and animated fictional characters, some of whom share their experience of a mental health problem and some of whom describe specific therapeutic techniques. Each module includes a written transcript and a subsequent written exercise that encourages the person to practice the described therapeutic technique. The program was adapted for this trial from a previous version of the intervention that was tailored to symptom profiles, with the tailoring removed because of a lack of increased efficacy [42]. The 12 modules of the current version of the program provide techniques predominantly based on cognitive behavioral therapy. The program included 8 core transdiagnostic modules (psychoeducation, getting help and support, cognitive reframing, problem-solving, mindfulness, managing relationships, exercise and diet, and sleep hygiene). It also contained 4 modules designed for specific mental health problems but focused on therapeutic targets broadly relevant to internalizing psychopathology, including 2 depression modules (behavioral activation and reducing rumination), 1 anxiety module (exposure), and 1 suicidality module (distress tolerance). The ordering of the program was self-directed, except that the participants were required to complete the psychoeducation module first. Participants had access to all 12 *FitMindKit* modules for 4 weeks.

Measures

Uptake was assessed as the number of participants who enrolled in the intervention and whether they completed any of the modules. *Reach* was assessed as the ratio of people who took up the intervention to those exposed to information or advertisements for the study within each pathway, based on estimates provided by the social media platform (Facebook and Instagram) and estimates of patient or customer throughput provided by the clinics and pharmacies. *Relative uptake* was examined by assessing the number of people taking up the

XSL•FO
RenderX

intervention within each pathway as a proportion of the reach of that pathway. Another indicator of reach, *inclusion*, was assessed on the basis of diversity in the characteristics of participants from each of the dissemination pathways; that is, no demographic groups were omitted from any of the pathways, and user characteristics in each pathway were largely consistent with each other and with the general population. Demographic characteristics included gender (male, female, or other gender), age (18-25, 26-35, 36-45, 46-55, 56-65, or >66 years), level of education (primary school, some secondary school or year 10 equivalents, year 12, Certificate Level I-IV, diploma or associate degree, bachelor's degree, graduate diploma or graduate certificate, master's degree, and doctoral degree), employment status (full-time, part-time or casual, unemployed, or not working because of study, maternity leave, or retirement), and language spoken at home (English only, English and another language, or another language only). Clinical characteristics included symptoms of depression (Patient Health Questionnaire-9 score, ranging from 0 to 27) [50], generalized anxiety (Generalized Anxiety Disorder-7 score, ranging from 0 to 21) [51], social anxiety (social anxiety disorder screener, range 0-16) [52], panic disorder (Panic Disorder Screener, range 0-16) [52], and suicidal thinking (Suicidal Ideation Attributes Scale, range 0-50) [53].

Costs of implementation were calculated on the basis of the costs of research staff time to engage each dissemination pathway and the costs of marketing (social media advertisements, posters, flyers, and tablets) associated with each pathway. The costs are provided in Australian dollars. *Adherence* was assessed as the number of modules of *FitMindKit* completed by the participants.

Data Analysis

Descriptive analyses were used to compare uptake, reach, and costs. Adherence (module completion) was compared between pathways using independent-sample 2-tailed *t* tests. The diversity of participants was assessed by comparing participant characteristics between the 3 pathways using chi-square statistics for categorical variables and the Welch test for continuous variables to account for homogeneity of variance between pathways. SPSS (version 26; IBM Corp) was used for statistical analyses.

Results

Uptake, Reach, and Inclusion

A total of 1014 participants were recruited into the *FitMindKit* program through three dissemination pathways: social media (991/1014, 97.73%), general practice (16/1014, 1.58%), and pharmacy (7/1014, 0.69%). A total of 1986 participants were recruited in the social media setting, and 991 (49.9%) were allocated to the control condition. Figure 1 shows that the rate of screening completion was highest in the social media pathway. Eligibility rates were similar across pathways; however, stark differences were seen in the rates of participants starting the intervention, with social media participants being considerably more likely to start the intervention.

The total potential exposure for the 3 pathways was assessed on the basis of the reach of social media advertising and estimates provided by the clinics and pharmacies of annual patient or customer numbers. Staff at the GP clinics estimated a total of 47,418 annual visits. On the basis of the Australian average of 5.6 annual visits per person [54], this represents 8468 individuals. Pharmacy visits were harder to estimate but based on the adult population of the catchment area (n=306,000) and the number of community pharmacies in the region (n=80), it was estimated that 14,345 adults visited 1 of the 5 pharmacies annually under the assumption that 75% of the adult population visited a community pharmacy. Reach of the social media advertisement was 98,135 individuals, with 448,797 impressions of the advertisements delivered.

On the basis of these estimates, 1 in every 50 people exposed to the social media advertising took up the intervention (991/49,068, 2.02% of individuals accounting for the control condition), compared with 1 in every 441 in GP clinics (16/7057, 0.22% of patients over 10 months) and 1 in every 1708 in pharmacies (7/11,953, 0.06% of customers over 10 months). Given the low rates of uptake in GP or pharmacy settings, we also examined the rates of screening completion as a proportion of exposed individuals as an alternative metric of reach. Screening completion rates were 4.91% (2408/49,068) in social media, 3.81% (269/7057) in GP clinics, and 1.81% (216/11,953) in pharmacies.

Table 1 presents the characteristics of the participants in each of the 3 dissemination pathways. There were no demographic groups that were poorly represented in any of the pathways, except there were few participants who spoke a language other than English, and there was an underrepresentation of men. When comparing the inclusion of demographic subgroups across the 3 pathways, there was some evidence that the social media sample trended toward being younger, more women, and more likely to speak only English. However, the only significant demographic difference was in employment status (χ^2_2=12.3; *P*=.002), indicating that participants in the social media pathway were more likely to be employed in work than those in the other pathways. For the GP and pharmacy pathways, all of those not working were classified as *not in the labor force* (eg, retired, student, and maternity leave) rather than unemployed, whereas in the social media pathway, 31.8% (69/217) of those not working were unemployed. In terms of mental health symptoms, there were significant differences in generalized anxiety symptoms, with GP participants having less severe anxiety than the other pathways (Welch test=7.00; *P*=.01), with a similar trend observed in panic symptoms. There were no significant differences in depression symptoms or suicidal ideation.

XSL•FO
RenderX

Table 1. Characteristics of participants in the 3 pathways (N=1014).

Characteristics	Social media (n=991)	GP[a] clinic (n=16)	Pharmacy (n=7)	Chi-square (*df*)	Welch test	*P* value
Gender (female), n (%)	841 (84.9)	11 (68.8)	6 (85.7)	3.1 (1)	N/A[b]	.21
Age group (years), n (%)				8.9 (2)	N/A	.06
18-35	413 (41.7)	4 (25)	2 (28.6)			
36-55	418 (42.2)	6 (37.5)	2 (28.6)			
≥56	160 (16.1)	6 (37.5)	3 (42.9)			
Educational attainment, n (%)				2.7 (1)	N/A	.26
Less than bachelor's degree	437 (44.1)	6 (37.5)	1 (14.3)			
Bachelor's or higher	558 (56.3)	10 (62.5)	6 (85.7)			
Employment, n (%)				12.3[c] (1)	N/A	.002[c]
Full-time or part-time	774 (78.1)	7 (43.8)	4 (57.1)			
Unemployed or not in labor force	217 (21.9)	9 (56.3)	3 (42.9)			
Language spoken at home, n (%)				5.01 (1)	N/A	.08
English only	902 (91)	12 (75)	6 (85.7)			
Other languages	89 (9)	4 (25)	1 (14.3)			
PHQ-9[d] depression score, mean (SD)	10.60 (4.95)	8.94 (5.77)	9.57 (3.46)	N/A	0.90	.36
GAD-7[e] anxiety score, mean (SD)	8.16 (4.38)	4.75 (3.54)	9.14 (4.56)	N/A	7.00[c]	.01[c]
PADIS[f] panic score, mean (SD)	3.20 (2.97)	1.63 (2.36)	4.14 (3.34)	N/A	3.57	.06
SOPHS[g] social anxiety score, mean (SD)	6.17 (3.80)	4.88 (3.28)	7.29 (1.89)	N/A	2.33	.14
SIDAS[h] suicidal ideation score, mean (SD)	6.47 (8.69)	4.69 (10.40)	6.14 (7.76)	N/A	1.80	.17

[a]GP: general practitioner.

[b]N/A: not applicable.

[c]$P<.05$.

[d]PHQ-9: Patient Health Questionnaire-9.

[e]GAD-7: Generalized Anxiety Disorder-7.

[f]PADIS: Panic Disorder Screener.

[g]SOPHS: Social Phobia Screener.

[h]SIDAS: Suicidal Ideation Attributes Scale.

Costs of Implementation

Social media advertising costs a total of US $8172; however, half of these advertising costs were directed toward recruiting participants for the control condition. Flyers and posters for clinics and pharmacies cost US $656, along with tablets (hardware and software) to enable screening in these settings cost US $1932.

Staff time for the social media advertising was estimated at 30 minutes per week to oversee advertising and billing for a total of US $734. Staff time for clinics and pharmacies was estimated at 60 minutes per site per week, which included monthly visits to each site to ensure the materials remained visible and the tablets were operational, as well as the travel time and time to maintain relationships and engagement with each site (both during and between site visits). After accounting for travel costs and staff time, these costs were US $7610 for the 5 GP clinics and US $7610 for the 5 pharmacies.

On the basis of the total costs of social media dissemination (US $4820), the cost of disseminating the intervention was US $4.87 per user. The cost of GP clinic dissemination (US $8904) was US $557 per user and that for pharmacy dissemination (US $8904) was US $1272 per user.

Adherence

Users in the social media pathway completed an average of 2.21 (SD 3.40) modules compared with those in the GP pathway (mean 3.56, SD 4.15) and in the pharmacy pathway (mean 1.14, SD 0.99). The 2-way comparisons between pathways were not significant (social media vs GP: $t_{1005}=1.57$, $P=.12$; social media vs pharmacy: $t_{996}=.83$, $P=.41$; GP vs pharmacy: $t_{21}=1.51$, $P=.15$), although high rates of type II errors were likely because of the small sample sizes.

XSL•FO
RenderX

Discussion

Principal Findings

This study compared 3 dissemination pathways for delivering a self-guided internet-based transdiagnostic intervention to reduce the symptoms of depression and anxiety. The findings supported hypotheses 1 and 2, indicating that delivering such interventions directly to end users through web-based marketing in social media leads to considerably greater uptake and reach than dissemination through primary care clinics and community pharmacies. Disseminating internet interventions to users through the same modality as the intervention (ie, social media dissemination) is likely to minimize behavioral gaps between being offered an intervention and engaging with the intervention. The disconnection between the clinic or pharmacy setting and an internet intervention may be an insurmountable gap unless the individual is highly motivated and supported to engage. Despite the low intervention uptake, the results suggest that screening for mental ill health in GP clinics and pharmacies appears to be feasible, with completion rates similar to those in the social media setting. By virtue of the low intervention uptake in GP and pharmacy settings, the per user cost of delivering the intervention was approximately 100 to 250 times greater in these 2 settings than in social media dissemination, supporting hypothesis 3. It appeared that GP users typically completed 1.5 modules more than social media users, although the difference was not significant by virtue of the small sample size. Contrary to hypothesis 4, users from the pharmacy pathway had lower adherence rates.

Exploratory analyses revealed variations in the characteristics of users across the 3 pathways. All pathways showed an underrepresentation of males, greater than would be expected based on the gender-specific prevalence of depression and anxiety [55]. Although web-based social media users tended to be younger (413/991, 41.7% vs 6/23, 26%), age distributions were not significantly different across pathways. GP and pharmacy participants were much more likely to be outside the labor force than the general population; in Australia, 66% of adults were employed, and 5.6% were unemployed [56], with the remainder not in the labor force because of reasons such as retirement or studying. This distribution most closely resembles that of the social media pathway, with working adults underrepresented in the other 2 pathways. Linguistic diversity was underrepresented [57] in all pathways, particularly on social media, which suggests that language or culture may be a barrier to engagement with internet interventions delivered in English.

Finally, there was some indication that anxiety symptoms were less severe among GP clinic participants than in other settings, with significantly lower generalized anxiety scores, although caution is needed because of the small sample size. Although screening for general psychological distress restricted the range of mental ill health in the sample, the findings suggest that participants engaged through GP clinics may have less severe anxiety than those who engaged through social media. Anxiety may be a distinct barrier to engagement with health services, such that recruitment in primary care may miss a sector of the population that experiences high levels of anxiety symptoms. Such differences were not observed for depression symptoms.

This study has several important implications for research and for the dissemination of internet-based interventions in the community. Clearly, for self-guided interventions, dissemination through web-based marketing appears to be considerably more efficient and cost-effective than dissemination through clinics or pharmacies. This may reflect the reluctance of people to engage with an intervention that is seen as distinct from the care they receive from the provider. It is possible that privacy concerns may influence decisions to engage more in public settings than when using the internet in private settings [58]. Furthermore, the web-based modality of the intervention may be more conducive to direct web-based dissemination than via the use of posters. Engaging directly with an intervention through a link may have fewer barriers to participation than requiring typing in a link from a poster, using a QR code from a poster, or using a shared device (noting that these data were collected before the COVID-19 pandemic). In contrast, engaging with a digital service advertised in a GP clinic or pharmacy may be seen by some as having implicit professional endorsement or greater credibility, although such effects were not observed here. Further research to understand the motivations of users across diverse dissemination settings may provide further insights into the barriers to and facilitators of implementation.

Greater success in dissemination in the web-based pathway did not result in significantly poorer adherence or a marked restriction in the diversity of participants. Although it may be slightly challenging to engage older adults or people from linguistically diverse backgrounds through social media marketing, there are methods for attracting specific subgroups through this pathway, such as tailored advertising and targeted platforms or groups. Nevertheless, the success of social media dissemination appears to outweigh these potential challenges, as the number of participants in all subgroups far exceeded those in the other pathways.

Limitations

There were some limitations to this study. The marketing materials were not consistent across the 3 pathways, with more information provided in the flyers or posters than in the social media advertising. It is possible that other marketing approaches in GP clinics and pharmacies may be more effective. However, we were limited by the constraints of these private, commercial enterprises in the types of marketing we could deliver, and there are few other methods that have been shown to be successful. Although informal testing of marketing materials was conducted before the study, formal testing of how the design of marketing materials influences willingness to participate would be a useful addition to future research. Although the catchment areas for all pathways were matched, the numbers exposed to marketing differed across pathways. Although we were able to estimate and account for these differences, there may have been some minor inaccuracies in these estimates. Furthermore, as noted previously, the way people engage with posters and tablets in a public setting may be different from how they engage with an advertising link in their social media accounts. It is likely that engagement would be considerably higher for interventions

XSL•FO
RenderX

that are directly recommended by a clinician rather than those from referral through marketing material within the clinic. Although this study used a passive engagement strategy that was not reliant on staff, it is possible that staff attitudes may have influenced engagement, although these attitudes were not directly assessed. Social media participants were recruited into a randomized controlled trial of the intervention, with the current data only reporting on those participants allocated to the active arm of the trial. Primary care and pharmacy participants were recruited directly to the web-based intervention, with no comparison arm. It is possible that barriers to engagement in trials may have contributed to differences in dissemination outcomes for the social media arm, although these barriers are unlikely to have improved outcomes. Cost estimates excluded the time of staff in clinics and pharmacies. As noted, the low numbers in the GP and pharmacy settings preclude firm conclusions about the differences in user characteristics and adherence. Finally, it is not clear whether these findings would generalize to clinician-guided interventions, non–internet-based interventions, or internet interventions for other health conditions. Involvement of providers in the referral process or characteristics of the health system (eg, costs and attitudes) may also influence engagement with interventions.

Conclusions

Given the low use of web-based mental health programs but the substantial evidence for their effectiveness, there is an urgent need to improve the uptake of web-based mental health interventions in the general community. This project has identified that a web-based dissemination pathway is the most efficient and cost-effective way of delivering self-guided internet-based mental health programs to people in the community without considerably sacrificing diversity of reach or adherence. Although screening in GP clinics and pharmacies appears to be feasible, in these settings, the behavior gap between completing an assessment and signing up for a digital intervention appears to be considerable.

Acknowledgments

This research was funded by a grant from the National Health and Medical Research Council (Australia) 1142363. The study is related to a larger research program funded by the European Union Horizons 2020 Project 733025, ImpleMentAll, although the study received no funding from the European Union. PJB and ALC were supported by National Health and Medical Research Council fellowships 1158707 and 1173146, respectively. LMF was supported by the Australian Research Council Discovery Early Career Research Award, DE190101382.

The authors thank the collaborators on the wider ImpleMentAll study and the general practitioner clinics and pharmacies who contributed to the dissemination activity.

Authors' Contributions

PJB and ALC designed the study and obtained project funding. PJB, AG, and EK drafted the manuscript. PJB conducted the analyses. EK and LMF contributed to the study management, data collection, and drafting of the manuscript. CV and JS contributed to the study management and oversaw the overarching ImpleMentAll study. All the authors critically edited and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Kessler RC, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
2. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018 Feb 12;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
3. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
4. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 2021 Nov 06;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
5. Breslau J, Lane M, Sampson N, Kessler RC. Mental disorders and subsequent educational attainment in a US national sample. J Psychiatr Res 2008 Jul;42(9):708-716 [FREE Full text] [doi: 10.1016/j.jpsychires.2008.01.016] [Medline: 18331741]
6. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. PLoS One 2014 Feb 28;9(2):e90311 [FREE Full text] [doi: 10.1371/journal.pone.0090311] [Medline: 24587318]

XSL•FO
RenderX

7. Kessler RC, Heeringa S, Lakoma MD, Petukhova M, Rupp AE, Schoenbaum M, et al. Individual and societal effects of mental disorders on earnings in the United States: results from the national comorbidity survey replication. Am J Psychiatry 2008 Jun;165(6):703-711 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08010126] [Medline: 18463104]
8. Burgess PM, Pirkis JE, Slade TN, Johnston AK, Meadows GN, Gunn JM. Service use for mental health problems: findings from the 2007 National Survey of Mental Health and Wellbeing. Aust N Z J Psychiatry 2009 Jul;43(7):615-623. [doi: 10.1080/00048670902970858] [Medline: 19530018]
9. Wang PS, Angermeyer M, Borges G, Bruffaerts R, Tat Chiu W, DE Girolamo G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):177-185 [FREE Full text] [Medline: 18188443]
10. Alonso J, Liu Z, Evans-Lacko S, Sadikova E, Sampson N, Chatterji S, WHO World Mental Health Survey Collaborators. Treatment gap for anxiety disorders is global: results of the World Mental Health Surveys in 21 countries. Depress Anxiety 2018 Mar;35(3):195-208 [FREE Full text] [doi: 10.1002/da.22711] [Medline: 29356216]
11. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. Psychol Med 2014 Apr;44(6):1303-1317 [FREE Full text] [doi: 10.1017/S0033291713001943] [Medline: 23931656]
12. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
13. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
14. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
15. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided Internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
16. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Sep;49(9):776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
17. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
18. Norton P. Transdiagnostic approaches to the understanding and treatment of anxiety and related disorders. J Anxiety Disord 2017 Mar;46:1-3. [doi: 10.1016/j.janxdis.2017.02.004] [Medline: 28249669]
19. 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] [Medline: 27712544]
20. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
21. 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] [Medline: 32865497]
22. Karyotaki E, Kemmeren L, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, et al. Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. Psychol Med 2018 Nov;48(15):2456-2466 [FREE Full text] [doi: 10.1017/S0033291718000648] [Medline: 29540243]
23. 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] [Medline: 33471111]
24. 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] [Medline: 30994877]
25. Batterham PJ, Calear AL. Preferences for Internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
26. Batterham PJ, Neil AL, Bennett K, Griffiths KM, Christensen H. Predictors of adherence among community users of a cognitive behavior therapy website. Patient Prefer Adherence 2008 Feb 02;2:97-105 [FREE Full text] [Medline: 19920949]

XSL•FO
RenderX

27. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
28. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003 Oct 11;362(9391):1225-1230. [doi: 10.1016/S0140-6736(03)14546-1] [Medline: 14568747]
29. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
30. Griffiths KM, Farrer L, Christensen H. The efficacy of Internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 07;192(S11):S4-11. [doi: 10.5694/j.1326-5377.2010.tb03685.x] [Medline: 20528707]
31. 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] [Medline: 29422409]
32. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. J R Soc Med 2011 Dec;104(12):510-520 [FREE Full text] [doi: 10.1258/jrsm.2011.110180] [Medline: 22179294]
33. Westfall JM, Mold J, Fagnan L. Practice-based research--"Blue Highways" on the NIH roadmap. JAMA 2007 Jan 24;297(4):403-406. [doi: 10.1001/jama.297.4.403] [Medline: 17244837]
34. Green LW, Ottoson JM, García C, Hiatt RA. Diffusion theory and knowledge dissemination, utilization, and integration in public health. Annu Rev Public Health 2009;30:151-174. [doi: 10.1146/annurev.publhealth.031308.100049] [Medline: 19705558]
35. Orman J, O'Dea B. e-Therapy in primary care mental health. Aust J Gen Pract 2018 Apr;47(4):168-172. [doi: 10.31128/AJGP-11-17-4410] [Medline: 29621855]
36. Ebert DD, Berking M, Cuijpers P, Lehr D, Pörtner M, Baumeister H. Increasing the acceptance of Internet-based mental health interventions in primary care patients with depressive symptoms. A randomized controlled trial. J Affect Disord 2015 May 01;176:9-17. [doi: 10.1016/j.jad.2015.01.056] [Medline: 25682378]
37. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, García Campayo J, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017 Feb 27;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
38. Fletcher S, Clarke J, Sanatkar S, Baldwin P, Gunn J, Zwar N, et al. Recruiting to a randomized controlled trial of a Web-based program for people with type 2 diabetes and depression: lessons learned at the intersection of e-mental health and primary care. J Med Internet Res 2019 May 24;21(5):e12793 [FREE Full text] [doi: 10.2196/12793] [Medline: 31127718]
39. Bertilsson E, Serhal S, Emmerton L, Bosnic-Anticevich S, Krass I, Bereznicki B, et al. Pharmacists experience of and perspectives about recruiting patients into a community pharmacy asthma service trial. Res Social Adm Pharm 2021 Mar;17(3):595-605. [doi: 10.1016/j.sapharm.2020.05.012] [Medline: 32448767]
40. Drozd F, Vaskinn L, Bergsund HB, Haga SM, Slinning K, Bjørkli CA. The implementation of Internet interventions for depression: a scoping review. J Med Internet Res 2016 Sep 08;18(9):e236 [FREE Full text] [doi: 10.2196/jmir.5670] [Medline: 27608548]
41. Bührmann L, Schuurmans J, Ruwaard J, Fleuren M, Etzelmüller A, Piera-Jiménez J, ImpleMentAll consortium. Tailored implementation of Internet-based cognitive behavioural therapy in the multinational context of the ImpleMentAll project: a study protocol for a stepped wedge cluster randomized trial. Trials 2020 Oct 28;21(1):893 [FREE Full text] [doi: 10.1186/s13063-020-04686-4] [Medline: 33115545]
42. Batterham PJ, Calear AL, Farrer L, McCallum SM, Cheng VW. FitMindKit: randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
43. Batterham PJ, Calear AL, Gulliver A, Farrer LM. Efficacy of a transdiagnostic, video-based online program for reducing depression, anxiety, and suicidal ideation in adults: protocol for a randomised controlled trial. Contemp Clin Trials Commun 2019 Jun;14:100341 [FREE Full text] [doi: 10.1016/j.conctc.2019.100341] [Medline: 30911696]
44. Batterham PJ, Calear AL, Farrer L, Gulliver A, Kurz E. Efficacy of a transdiagnostic self-help Internet intervention for reducing depression, anxiety, and suicidal ideation in adults: randomized controlled trial. J Med Internet Res 2021 Jan 22;23(1):e22698 [FREE Full text] [doi: 10.2196/22698] [Medline: 33480860]
45. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
46. Batterham PJ, Calear AL, Sunderland M, Kay-Lambkin F, Farrer LM, Christensen H, et al. A brief intervention to increase uptake and adherence of an Internet-based program for depression and anxiety (enhancing engagement with psychosocial interventions): randomized controlled trial. J Med Internet Res 2021 Jul 27;23(7):e23029 [FREE Full text] [doi: 10.2196/23029] [Medline: 34313595]

XSL•FO
RenderX

47. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
48. Batterham PJ, Sunderland M, Carragher N, Calear AL, Mackinnon AJ, Slade T. The Distress Questionnaire-5: population screener for psychological distress was more accurate than the K6/K10. J Clin Epidemiol 2016 Mar;71:35-42. [doi: 10.1016/j.jclinepi.2015.10.005] [Medline: 26464194]
49. Batterham PJ, Sunderland M, Slade T, Calear AL, Carragher N. Assessing distress in the community: psychometric properties and crosswalk comparison of eight measures of psychological distress. Psychol Med 2018 Jun;48(8):1316-1324. [doi: 10.1017/S0033291717002835] [Medline: 28967345]
50. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. JAMA 1999 Nov 10;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
51. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
52. Sunderland M, Batterham PJ, Calear AL, Carragher N. The development and validation of static and adaptive screeners to measure the severity of panic disorder, social anxiety disorder, and obsessive compulsive disorder. Int J Methods Psychiatr Res 2017 Dec;26(4):e1561 [FREE Full text] [doi: 10.1002/mpr.1561] [Medline: 28370638]
53. van Spijker BA, Batterham PJ, Calear AL, Farrer L, Christensen H, Reynolds J, et al. The suicidal ideation attributes scale (SIDAS): community-based validation study of a new scale for the measurement of suicidal ideation. Suicide Life Threat Behav 2014 Aug;44(4):408-419. [doi: 10.1111/sltb.12084] [Medline: 24612048]
54. Frequent GP attenders and their use of health services in 2012–13. Australian Institute of Health and Welfare, Australian Government. 2015 Mar 19. URL: https://www.aihw.gov.au/reports/primary-health-care/frequent-gp-attenders-use-health-services-2012-13/contents/summary [accessed 2022-04-21]
55. Slade T, Johnston A, Oakley Browne MA, Andrews G, Whiteford H. 2007 National Survey of Mental Health and Wellbeing: methods and key findings. Aust N Z J Psychiatry 2009 Jul;43(7):594-605. [doi: 10.1080/00048670902970882] [Medline: 19530016]
56. Labour Force, Australia: headline estimates of employment, unemployment, underemployment, participation and hours worked from the monthly Labour Force Survey, March 2022. Australian Bureau of Statistics. 2021. URL: https://www.abs.gov.au/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release [accessed 2022-04-21]
57. The 2016 Census of Population and Housing. Australian Bureau of Statistics. Canberra, Australia: Australian Bureau of Statistics; 2016. URL: https://www.abs.gov.au/ausstats/abs@.nsf/lookup/by%20subject/1001.0~2016-17~main%20features~the%202016%20census%20of%20population%20and%20housing~10009 [accessed 2022-04-21]
58. MacDonnell K, Cowen E, Cunningham DJ, Ritterband L, Ingersoll K. Online recruitment of a non-help-seeking sample for an internet intervention: lessons learned in an alcohol-exposed pregnancy risk reduction study. Internet Interv 2019 Sep;17:100240 [FREE Full text] [doi: 10.1016/j.invent.2019.100240] [Medline: 30963031]

Abbreviations

DQ5: Distress Questionnaire 5
GP: general practitioner

Edited by R Kukafka; submitted 08.11.21; peer-reviewed by C Oehler, S Moritz; comments to author 05.01.22; revised version received 27.01.22; accepted 04.04.22; published 06.05.22.

Please cite as:
Batterham PJ, Gulliver A, Kurz E, Farrer LM, Vis C, Schuurmans J, Calear AL
The Effect of Dissemination Pathways on Uptake and Relative Costs for a Transdiagnostic, Self-guided Internet Intervention for Reducing Depression, Anxiety, and Suicidal Ideation: Comparative Implementation Study
J Med Internet Res 2022;24(5):e34769
URL: https://www.jmir.org/2022/5/e34769
doi:10.2196/34769
PMID:35522458

©Philip J Batterham, Amelia Gulliver, Ella Kurz, Louise M Farrer, Christiaan Vis, Josien Schuurmans, Alison L Calear. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2022. 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

XSL•FO
RenderX

unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.

XSL•FO
RenderX

Original Paper

Factors Predicting Engagement of Older Adults With a Coach-Supported eHealth Intervention Promoting Lifestyle Change and Associations Between Engagement and Changes in Cardiovascular and Dementia Risk: Secondary Analysis of an 18-Month Multinational Randomized Controlled Trial

Nicola Coley[1,2], PhD; Laurine Andre[1], MD; Marieke P Hoevenaar-Blom[3,4,5], PhD; Tiia Ngandu[6,7], MD, PhD; Cathrien Beishuizen[8], MD, PhD; Mariagnese Barbera[9,10], PhD; Lennard van Wanrooij[4], PhD; Miia Kivipelto[7,9,10,11], MD, PhD; Hilkka Soininen[9,12], MD, PhD; Willem van Gool[5], MD, PhD; Carol Brayne[13], MD; Eric Moll van Charante[5,8], MD, PhD; Edo Richard[3,5], MD, PhD; Sandrine Andrieu[1,2], MD, PhD; HATICE study group[14]; PRODEMOS study group[14]

[1]Center for Epidemiology and Research in Population health (CERPOP), University of Toulouse III Paul Sabatier (UPS), National Institute of Health and Medical Research (INSERM) mixed research unit (UMR) 1295, Toulouse, France

[2]Department of Epidemiology and Public Health, Toulouse University Hospital, Toulouse, France

[3]Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, Netherlands

[4]Department of Neurology, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

[5]Department of Public and Occupational Health, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands

[6]Population Health Unit, Finnish Institute for Health and Welfare, Helsinki, Finland

[7]Division of Clinical Geriatrics, Center for Alzheimer Research, Care Sciences and Society (NVS), Karolinska Institutet, Stockholm, Sweden

[8]Department of General Practice, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands

[9]Institute of Clinical Medicine, Department of Neurology, University of Eastern Finland, Kuopio, Finland

[10]Ageing Epidemiology Research Unit, School of Public Health, Imperial College London, London, United Kingdom

[11]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

[12]Neurocenter, Neurology, Kuopio University Hospital, Kuopio, Finland

[13]Cambridge Public Health, University of Cambridge, Cambridge, United Kingdom

[14]See Acknowledgments

Corresponding Author:
Nicola Coley, PhD
Center for Epidemiology and Research in Population health (CERPOP)
University of Toulouse III Paul Sabatier (UPS), National Institute of Health and Medical Research (INSERM) mixed research unit (UMR) 1295
37 allées Jules Guesde
Toulouse, 31000
France
Phone: 33 561145680
Email: nicola.coley@inserm.fr

Abstract

Background: Digital health interventions could help to prevent age-related diseases, but little is known about how older adults engage with such interventions, especially in the long term, or whether engagement is associated with changes in clinical, behavioral, or biological outcomes in this population. Disparities in engagement levels with digital health interventions may exist among older people and be associated with health inequalities.

Objective: This study aimed to describe older adults' engagement with an eHealth intervention, identify factors associated with engagement, and examine associations between engagement and changes in cardiovascular and dementia risk factors (blood pressure, cholesterol, BMI, physical activity, diet, and cardiovascular and dementia risk scores).

XSL•FO
RenderX

Methods: This was a secondary analysis of the 18-month randomized controlled Healthy Ageing Through Internet Counselling in the Elderly trial of a tailored internet-based intervention encouraging behavior changes, with remote support from a lifestyle coach, to reduce cardiovascular and cognitive decline risk in 2724 individuals aged ≥65 years, recruited offline in the Netherlands, Finland, and France. Engagement was assessed via log-in frequency, number of lifestyle goals set, measurements entered and messages sent to coaches, and percentage of education materials read. Clinical and biological data were collected during in-person visits at baseline and 18 months. Lifestyle data were self-reported on a web-based platform.

Results: Of the 1389 intervention group participants, 1194 (85.96%) sent at least one message. They logged in a median of 29 times, and set a median of 1 goal. Higher engagement was associated with significantly greater improvement in biological and behavioral risk factors, with evidence of a dose-response effect. Compared with the control group, the adjusted mean difference (95% CI) in 18-month change in the primary outcome, a composite z-score comprising blood pressure, BMI, and cholesterol, was −0.08 (−0.12 to −0.03), −0.04 (−0.08 to 0.00), and 0.00 (−0.08 to 0.08) in the high, moderate, and low engagement groups, respectively. Low engagers showed no improvement in any outcome measures compared with the control group. Participants not using a computer regularly before the study engaged much less with the intervention than those using a computer up to 7 (adjusted odds ratio 5.39, 95% CI 2.66-10.95) or ≥7 hours per week (adjusted odds ratio 6.58, 95% CI 3.21-13.49). Those already working on or with short-term plans for lifestyle improvement at baseline, and with better cognition, engaged more.

Conclusions: Greater engagement with an eHealth lifestyle intervention was associated with greater improvement in risk factors in older adults. However, those with limited computer experience, who tended to have a lower level of education, or who had poorer cognition engaged less. Additional support or forms of intervention delivery for such individuals could help minimize potential health inequalities associated with the use of digital health interventions in older people.

(J Med Internet Res 2022;24(5):e32006) doi:10.2196/32006

KEYWORDS

aging; eHealth; disparities; engagement; prevention; cardiovascular; lifestyle; risk factors

Introduction

Background

The number of people aged ≥60 years increased from 382 million worldwide in 1980 to 962 million in 2017 and is expected to reach nearly 2.1 billion by 2050 [1]. In parallel, there are an increasing number of cases of age-related diseases, including cardiovascular disease (CVD) and dementia, placing an ever-increasing burden on health and social care systems [2]. For example, the worldwide cost of dementia increased by 35% between 2010 and 2015 to reach US $818 billion and was estimated to exceed US $1 trillion in 2018 [3]. CVD and dementia share many potentially modifiable lifestyle-based risk factors, including physical inactivity, unhealthy diet, obesity, and hypertension [4,5], offering opportunities for prevention that could bring about huge public health gains. There is a need to establish, using rigorously conducted research studies, the extent to which interventions might influence behavior in the short, medium, and longer term, providing evidence for the best policies to reduce noncommunicable disease risk.

Digital health tools are a possible approach for delivering preventive interventions for CVD and dementia [6,7], which, if effective and efficient, can be rolled out at scale. However, nonuse of digital health interventions is a fundamental problem, with persistent reports of high discontinuation rates, even in research studies involving atypically motivated individuals [8,9]. Specifying the *dose* of nonpharmacological interventions, in terms of engagement, is inherently more difficult than for drug treatments, even more so for digital interventions where use is often at participants' discretion [8] and requires more investment and motivation than simply taking a daily medication [10]. Although increased engagement with digital interventions is associated with greater improvements in health outcomes, including behavior change, in young and middle-aged adults [10-12], very little is known about engagement with digital interventions [13], or its association with health outcomes, in older people, especially in the long term.

A concern about using digital health interventions in older populations is that they may further widen existing health inequalities [14]. Although the use of digital technologies is increasing in this age group, *digital exclusion* is still common, particularly in individuals aged >75 years, and older adults who do not use the internet have poorer health and lower socioeconomic status than those who do [15]. Even among older internet users, the levels of engagement with digital technologies may vary and be associated with individuals' characteristics. It is vital to understand better how older people use and interact with such tools and to identify potential disparities in use.

Objectives

To explore this, we drew on data from a large international trial of an eHealth intervention designed to encourage behavior changes for the prevention of CVD and cognitive decline in older individuals to (1) describe engagement with the different components of the eHealth intervention, (2) identify factors associated with engagement, and (3) examine associations between engagement and changes in cardiovascular and dementia risk factors.

Methods

Setting and Participants

We analyzed data from the previously described 18-month Healthy Ageing Through Internet Counselling in the Elderly (HATICE) parallel group randomized controlled trial (ISRCTN48151589) [16-18]. Between March 2015 and August

XSL•FO
RenderX

2016, a total of 2724 dementia-free community dwellers aged ≥65 years with at least basic computer literacy and either 2 or more CVD risk factors (hypertension, dyslipidemia, overweight, smoking, or physical inactivity) or a history of CVD or diabetes were enrolled in Finland, France, and the Netherlands. Participants were randomized in a 1:1 ratio to either (1) the intervention, a multicomponent internet-based platform designed to encourage lifestyle changes, with remote support from a lifestyle coach (see the *Intervention* section), or (2) a control group receiving access to a simple static internet platform containing only basic health information and no coach support. Participants were recruited offline, primarily through a population registry (Finland), commercial mailing lists and a prevention center (France), and general practitioners (Netherlands) [17,19]. The intervention had a modest but significant beneficial effect on the trial's primary outcome, a composite cardiovascular risk score [17]. Clinical, demographic, and biological data were collected during face-to-face study visits at baseline and 18 months. Data concerning lifestyle, mood, and health self-management were self-reported via the study's web-based platform at baseline and 12 and 24 months. Adverse events were self-reported on the web-based platform every 3 months.

Ethics Approval

The local ethical committees in each country approved the protocol (Academic Medical Centre, the Netherlands: METC 2014_126; Northern Savonia Hospital District Research Ethics Committee, Finland: 35/2014; Comité de Protection des Personnes Sud Ouest et Outre Mer, France: 2014-A01287–40), and all participants provided written informed consent.

Intervention

Intervention group participants had access to a secure internet-based platform (Figure 1), with remote support from a lifestyle coach trained in motivational interviewing and healthy lifestyle advice. Full details of the development and content of the platform have been previously published [18]. The intervention aimed to facilitate the self-management of cardiovascular risk factors, including hypertension, obesity, physical inactivity, diet, smoking, diabetes, and hypercholesterolemia, to improve the overall risk profile. It was designed using national and European guidelines for primary and secondary CVD prevention [20] and input from members of the target population, health professionals, and patient organizations [16,21,22].

After secure login, participants were able to (1) view their individual cardiovascular risk profile (based on baseline measurements), (2) set personal goals for lifestyle change and make corresponding action plans, (3) monitor goals by entering data (eg, blood pressure measurements or food diaries) and receive graphical or automated feedback, and (4) obtain health information from education modules (including text, videos, and quizzes) and peer-to-peer videos. News items related to CVD, healthy aging, or eHealth were added regularly to the platform. All content was provided in the local language, and advice was adapted to local guidelines where necessary. Owing to the older age of the study's participants and their expected level of computer experience, the navigation structure and layout of the intervention platform were kept as simple as possible (Figure 1).

Coaches met with participants face-to-face at baseline and thereafter communicated with them via a computer messaging system. There was also a booster telephone call at 12 months. Using motivational interviewing techniques, they supported participants in making lifestyle changes by encouraging them to prioritize up to 3 health factors (the home page layout then reflected the chosen health priorities) and set at least one goal at baseline, interact with the platform, and set additional goals over time. Coaches also provided motivational feedback. Participants could see a photograph and the name of their coach when they logged in to the intervention platform (Figure 1).

The intervention was designed by the academic researchers involved in the project. Technical development was performed by Vital Health Software.

XSL•FO
RenderX

Figure 1. Screenshots of the HATICE intervention platform: (A) home page and (B) measurements page. HATICE: Healthy Ageing Through Internet Counselling in the Elderly.



Outcomes

Engagement Outcomes

Engagement with digital health interventions, similar to engagement with serious games [23], is thought to encompass behavioral, emotional, and cognitive factors [24-26], but there is no consensus on how it should be measured [10]. Similar to previous studies [10,11,27,28], we assessed engagement (only in the intervention group) with our eHealth intervention through system use metrics, including number and dates of logins, number of goals set, messages sent to coaches, monitoring measurements or goal diary entries, and percentage of advice and education materials read.

Different studies have found different components of eHealth intervention use, including number of logins, number of activities completed per login, percentage of study modules

XSL•FO
RenderX

completed, amount of goals set, and number of self-monitoring measures, to be associated with outcomes [26,29,30], and it has been suggested that composite measures may be the best way to measure engagement or adherence with such interventions [10]. Therefore, we developed a composite indicator of overall engagement throughout the study, defined as the sum of the points obtained for logins (0, 1, and 2 points for the first, second, and third tertiles of total logins, respectively), number of goals set (0: 0 points; 1: 1 point; ≥2: 2 points), sending of at least one message to the coach (no: 0 points; yes: 1 point), entering at least one measurement (no: 0 points; yes: 1 point), and reading some of the advice and education materials (no: 0 points; yes: 1 point). The composite engagement score ranged from 0 to 7 points, with higher scores indicating greater engagement. It was categorized into low (0-2 points), moderate (3-5 points), and high (6-7 points) engagement, thus avoiding potential difficulties with nonlinear relationships with outcome measures [26,29]. This categorization was decided before the analysis and was intended to capture *meaningful* differences in engagement with the platform use among the 3 groups (ie, *low*, *moderate*, and *high* engagers). In sensitivity analyses, we categorized the engagement scores using tertiles.

In addition, login data were used to study platform use over time. As participants were asked to log in every 3 months to complete an adverse event questionnaire, and at 12 and 18 months to complete study evaluations, and we could not distinguish these logins from other types, rather than calculating the time to last login, we calculated the time to the first occurrence of nonuse attrition, that is, no login during the previous month [8]. Because they only logged in once every 3 months, at most, participants who only logged in to complete adverse event questionnaires or study evaluations (and did not otherwise use the intervention platform) were considered to display nonuse attrition.

Risk Factor Outcomes

The trial's primary outcome was a composite cardiovascular risk *z* score based on systolic blood pressure, low-density lipoprotein cholesterol, and BMI [17]. Secondary outcomes included individual components of the composite *z* score, physical activity [31], dietary intake (Mediterranean Diet Adherence Screener score) [32], and estimated cardiovascular (Systematic Coronary Risk Estimation–Older People) [33] and dementia (Cardiovascular Risk Factors, Aging, and Incidence of Dementia score) [34] risk.

Predictor Variables

The following baseline variables were assessed as predictors of engagement and nonuse attrition: age, sex, level of education, country of residence, living status, history of CVD or diabetes, current smoking, physical activity, hypertension, dyslipidemia, obesity, intention to make lifestyle changes, cognition, depressive symptoms, anxiety, chronic condition self-management, physical performance, computer use during the 4 weeks before baseline, and diet. Table S1 in Multimedia Appendix 1 provides further details.

Statistical Analyses

Baseline characteristics were described using means (SD), medians (IQR), or numbers (%) and were compared among low, moderate, and high engagers using 1-way ANOVA or Kruskal-Wallis tests for continuous variables and chi-square tests for categorical variables.

Baseline predictors of overall engagement (categorized as low, moderate, or high) were assessed using a multivariate generalized ordered logit (partial proportional odds) model using the Stata *gologit2* command (StataCorp LP) [35]. In this model, the proportional odds assumption (ie, that the relationship between each pair of outcome categories was the same) was assessed using Wald tests for each predictor variable. If the assumption held, only 1 coefficient was calculated for the predictor variable, as in standard ordinal logistic regression. If it was violated, separate coefficients were calculated for the comparison of low versus moderate and high engagement categories and for low and moderate versus high engagement categories. The initial multivariate model included all variables associated with engagement in bivariate models at the .2 significance level, and the final model was determined using a manual backward stepwise selection procedure (sequentially eliminating variables with a $P>.05$).

Baseline factors associated with the first occurrence of nonuse attrition were examined using multivariate Cox proportional hazards models. As the proportional hazards assumption (verified using Schoenfeld residuals) was not met in the initial analysis, the follow-up period was split, following visual inspection of Kaplan-Meier survival curves, into early (ie, months 0-2) and late (ie, month 3 onward) periods, and analyses were run separately for each period. Variable selection was performed as for the generalized ordered logit model.

Finally, 18-month changes in the HATICE primary and secondary outcome variables were compared between the control group and the 3 engagement categories in the intervention group using linear regression models, adjusted for baseline variables associated with engagement, age, education, physical status, and smoking. Models were further adjusted for the baseline score of the outcome of interest if it differed significantly among engagement groups.

All analyses were exploratory and performed using Stata (version 14.1).

Results

Description of Engagement

The median number of logins per participant in the intervention group (N=1389) during the 18-month follow-up period was 29 (IQR 16-48; range 0-700; Figure S1A in Multimedia Appendix 1). In comparison, the median number of logins to the static platform in the control group (N=1335) was 12 (IQR 9-16).

Of the 1389, intervention group participants, 1194 (85.96%) sent at least one message to their coach during the 18-month study period (Figure S1B in Multimedia Appendix 1), and the median (IQR) number of messages sent was 6 (2-10). The median (IQR) number of goals set was 1 (1-2), and of the 1389

XSL•FO
RenderX

participants, 151 (10.87%) did not set any goals, and 560 (40.32%) set 2 or more goals (Figure S1C in Multimedia Appendix 1). Weight was the health factor most commonly targeted by goals, followed by physical activity, and nutrition (Figure S1D in Multimedia Appendix 1). Participants were most likely to read advice and education pages for cholesterol, blood pressure, and diabetes (Figure S1E in Multimedia Appendix 1). Physical activity measurements (min/week, subjectively reported) were the most frequent type of monitoring data entered, followed by weight and blood pressure measurements (Figure S1F in Multimedia Appendix 1). Additional descriptive engagement data are provided in Table S2 in Multimedia Appendix 1.

The median (IQR) composite engagement score was 5 (3-6), and of the 1389 participants, 208 (14.97%) were classified as having low engagement with the platform, 681 (49.03%) moderate engagement, and 500 (36%) high engagement (Table S3 in Multimedia Appendix 1). All components of platform use significantly increased across the three categories (and tertiles; Table S4 in Multimedia Appendix 1).

Predictors of Engagement

At baseline, participants in the low engagement category were younger, had a lower level of education, had poorer cognitive and physical performance, and had more depressive symptoms than those who engaged more. They were also more often from the Netherlands, more likely to be smokers, and less likely to have used a computer in the preceding 4 weeks or be planning or already acting on lifestyle change (Table 1).

In the multivariate analysis (Table 2), increasing engagement was independently predicted by country of residence, having short-term (ie, within the next month) plans for lifestyle change or acting on it for more than 6 months, and regular computer use at baseline. Furthermore, compared with those in the low engagement category, participants in the moderate and high engagement categories had better baseline cognitive performance, and compared with those in the low and moderate categories, those in the high engagement category were more likely to be women.

XSL•FO
RenderX

Table 1. HATICE[a] participants' (intervention group) baseline characteristics by overall engagement during the trial.

	Low engagement (N=208)	Moderate engagement (N=681)	High engagement (N=500)	*P* value
Age (years), median (IQR)	69.6 (67.5-73.9)	69.5 (67.3-72.8)	69.0 (67.0-72.4)	.05
Men, n (%)	102 (49)	385 (56.5)	244 (48.8)	.02
Education[b], n (%)				.02
Low	82 (39.4)	201 (29.5)	134 (26.8)	
Medium	58 (27.9)	206 (30.3)	159 (31.8)	
High	68 (32.7)	274 (40.2)	207 (41.4)	
Country, n (%)				.006
Netherlands	121 (58.2)	402 (59.0)	244 (48.8)	
France	22 (10.6)	86 (12.6)	69 (13.8)	
Finland	65 (31.3)	193 (28.3)	187 (37.4)	
Living with partner, n (%)	146 (70.2)	500 (73.4)	361 (72.2)	.65
Cognitive *z* score[c], mean (SD)	−0.14 (0.69)	0.03 (0.58)	0.00 (0.08)	.003
SPPB[d,e]<10, n (%)	47 (22.6)	93 (13.7)	78 (15.6)	.008
Depressive symptoms[f], n (%)	27 (13)	54 (7.9)	39 (7.8)	.05
HADS[g] anxiety score[h], median (IQR)	4 (2-6)	4 (2-6)	4 (2-6)	.16
History of CVD[i], n (%)	60 (28.9)	210 (31)	154 (30.9)	.83
Diabetes, n (%)	41 (19.7)	154 (22.7)	101 (20.2)	.50
Hypertension, n (%)	170 (82.9)	557 (83.5)	409 (83.3)	.98
Dyslipidemia, n (%)	201 (97.1)	653 (96)	480 (96.6)	.74
Currently smoking, n (%)	20 (12.1)	47 (7.5)	29 (5.9)	.03
Physically active[j], n (%)	128 (61.5)	452 (66.5)	334 (66.8)	.36
Obese, n (%)	82 (39.4)	262 (38.5)	185 (37)	.79
MEDAS[k] score[l], mean (SD)	5.8 (2.0)	6.2 (2.0)	6.0 (1.9)	.10
PIH[m] score[n], median (IQR)	87 (79-91)	87 (81-91)	86 (81-91)	.53
Trying to change lifestyle? n (%)				.001
No plans	25 (12)	45 (6.6)	30 (6)	
Long-term plans	27 (13)	78 (11.5)	40 (8)	
Short-term plans	35 (16.8)	82 (12)	88 (17.6)	
Short-term acting	39 (18.8)	120 (17.6)	92 (18.4)	
Long-term acting	82 (39.4)	356 (52.3)	250 (50)	
Computer use in the last 4 weeks, n (%)				<.001
No	22 (11)	18 (3)	3 (1)	
Yes, <7 hours/week	113 (55)	395 (58)	276 (55)	
Yes, ≥7 hours/week	72 (35)	267 (39)	221 (44)	

[a]HATICE: Healthy Ageing Through Internet Counselling in the Elderly.

[b]Low, medium, and high education levels correspond to basic, postsecondary nontertiary, and tertiary levels, respectively.

[c]Cognitive *z* score indicates average *z* scores of the Mini Mental Status Examination, Category Fluency, Stroop Color-Word Test, and Rey Auditory Verbal Learning Test.

[d]SPPB: Short Physical Performance Battery.

[e]Range 0-12 points, where higher scores indicate better performance.

[f]Geriatric Depression Scale–15 score ≤5.

XSL•FO
RenderX

[g]HADS: Hospital Anxiety and Depression Scale.

[h]Range 0-21, where higher scores indicate increasing symptoms of anxiety.

[i]CVD: cardiovascular disease.

[j]Defined as meeting the World Health Organization guidelines of ≥150 minutes' moderate-intensity or ≥75 minutes' vigorous-intensity physical activity per week.

[k]MEDAS: Mediterranean Diet Adherence Screener.

[l]Range 0-14, where higher scores indicate higher adherence to Mediterranean diet.

[m]PIH: Partners in Health.

[n]Range 0-96, where higher scores indicate better chronic disease self-management.

Table 2. Final multivariate[a] generalized ordered logistic regression model showing factors significantly associated with increasing overall engagement during follow-up (categorized as low, moderate, or high platform engagement; N=1238).

	OR[b] (95% CI)	*P* value[c]
Variables meeting the proportional odds assumption[d]		
Country		*.02*
Netherlands (ref[e])	1	N/A[f]
France	1.41 (0.98-2.02)	.07
Finland	1.55 (1.16-2.06)	.003
Trying to change lifestyle?		*.002*
No plans (ref)	1	N/A
Long-term plans	1.20 (0.70-2.07)	.51
Short-term plans	2.25 (1.33-3.80)	.002
Short-term acting	1.51 (0.92-2.50)	.11
Long-term acting	2.02 (1.26-3.25)	.004
Computer use in last 4 weeks before baseline visit		*<.001*
None (ref)	1	N/A
<7 hours/week	5.39 (2.66-10.95)	<.001
≥7 hours/week	6.58 (3.21-13.49)	<.001
Variables not meeting the proportional odds assumption[g]		
Low engagement (ref) vs moderate and high engagement		
Sex (male)	1.20 (0.84-1.72)	.31
Cognitive *z* score	1.67 (1.26-2.21)	<.001
Low and moderate engagement (ref) vs high engagement		
Sex (male)	0.77 (0.60-0.98)	.03
Cognitive *z* score	0.99 (0.81-1.22)	.95

[a]The following baseline variables were included in the initial multivariate model but did not remain significantly associated with engagement following a backward stepwise selection procedure: age, education, current smoking, physical status (Short Physical Performance Battery), depressive symptoms (Geriatric Depression Scale), anxiety (Hospital Anxiety and Depression Scale), and nutrition score.

[b]OR: odds ratio.

[c]*P* values in italics are overall Wald tests for categorical variables.

[d]For independent variables meeting the proportional odds assumption, the relationship between each pair of outcome categories (ie, moderate and high engagement vs low engagement and high engagement vs low and moderate engagement) is the same; therefore, only 1 OR is calculated per variable.

[e]ref: reference.

[f]N/A: not applicable.

[g]For independent variables not meeting the proportional odds assumption, separate ORs are calculated between each pair of outcome categories.

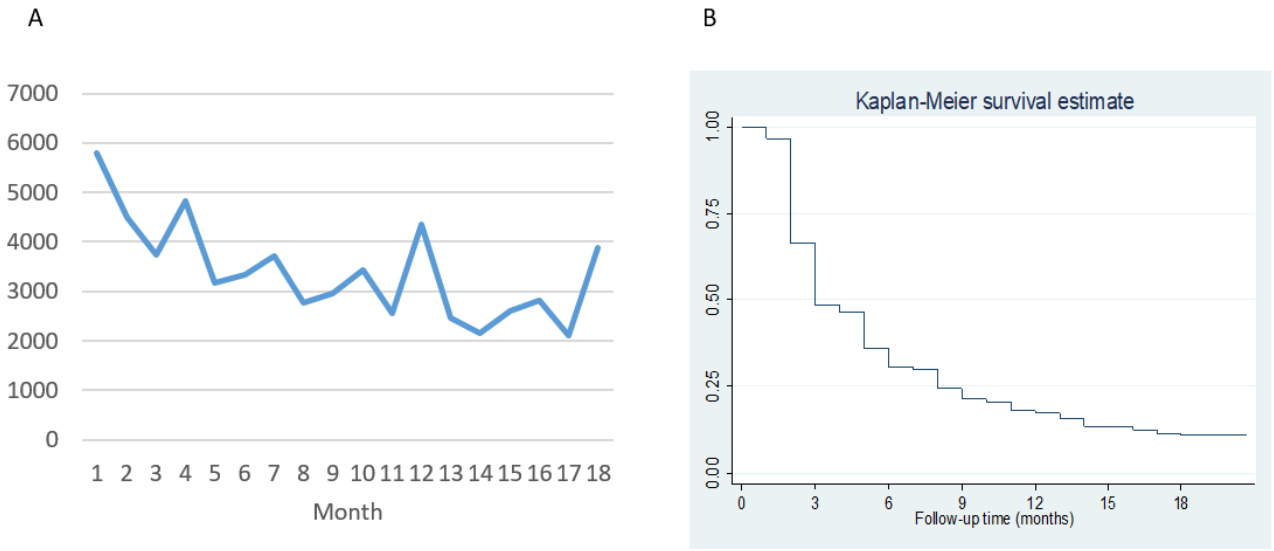
XSL•FO
RenderX

Changes Over Time in Engagement and Its Associated Factors

As shown in Figure 2A, intervention use (measured using logins) declined over time. The roughly (reverse) sigmoidal Kaplan-Meier nonuse attrition curve (Figure 2B) shows a sharp decrease in the proportion of participants logging in at least once during the previous month from the end of month 2 onward. In a sensitivity analysis using a 6-week nonuse attrition definition, the curve was similar, although slightly elongated (Figure S2 in Multimedia Appendix 1). The median time to the first occurrence of nonuse attrition was 3 months. Of the 1389 participants in the intervention group, 465 (33.48%) demonstrated early nonuse attrition (during months 1-2), 747 (53.78%) demonstrated late nonuse attrition (between 2 and 18 months), and 145 (10.44%) were highly consistent platform users, logging in at least once every month for the entire follow-up period (32/1389, 2.30% logged in every month, during which they participated in the trial but dropped out before the end of follow-up). The highly consistent users demonstrated significantly higher engagement with all parts of the platform than the other 2 groups, but even in the early nonuse attrition group, 24.30% (113/465) logged in to the platform at least once a month for at least 12 of the 18 months of follow-up (Table S5 in Multimedia Appendix 1).

Not using a computer regularly before baseline and a lower baseline chronic condition self-management score predicted early nonuse attrition, whereas living in the Netherlands and acting on lifestyle change for <6 months at baseline predicted late nonuse attrition (Table 3).

Figure 2. Changes in engagement over time in the intervention group: (A) total number of logins per month in the intervention group and (B) time to nonuse attrition (ie, no login during the previous month).



XSL•FO
RenderX

Table 3. Baseline factors associated with early (model 1) and late (model 2) nonuse attrition.

	HR[a]	95% CI	*P* value
Model 1: early nonuse attrition[b] (N=1351; 448 events)			
Computer use in the last 4 weeks			
None (ref[c])	1	N/A[d]	N/A
<7 hours/week	0.46	0.31-0.69	<.001
≥7 hours/week	0.44	0.29-0.66	<.001
Partners in Health score (points)[e]	0.99	0.98-1.00	.03
Model 2: late nonuse attrition[f] (N=848; 693 events)			
Country			
Netherlands (ref)	1	N/A	N/A
France	0.66	0.51-0.84	.001
Finland	0.57	0.47-0.69	<.001
Trying to change lifestyle?			
No plans to change lifestyle (ref)	1	N/A	N/A
Long-term plans to change lifestyle	0.96	0.66-1.40	.83
Short-term plans to change lifestyle	1.31	0.96-1.78	.09
Short-term acting on lifestyle change	1.49	1.15-1.93	.002
Long-term acting on lifestyle change	1.15	0.91-1.45	.23

[a]HR: hazard ratio.

[b]The first instance of nonuse attrition during months 1 to 2. The following baseline variables were included in the initial multivariate model but did not remain significantly associated with early nonuse attrition following a backward stepwise selection procedure: education, history of cardiovascular disease, history of diabetes, history of hypertension, Geriatric Depression Scale score, and verbal fluency score.

[c]ref: reference.

[d]N/A: not applicable.

[e]Higher scores indicate better chronic disease self-management.

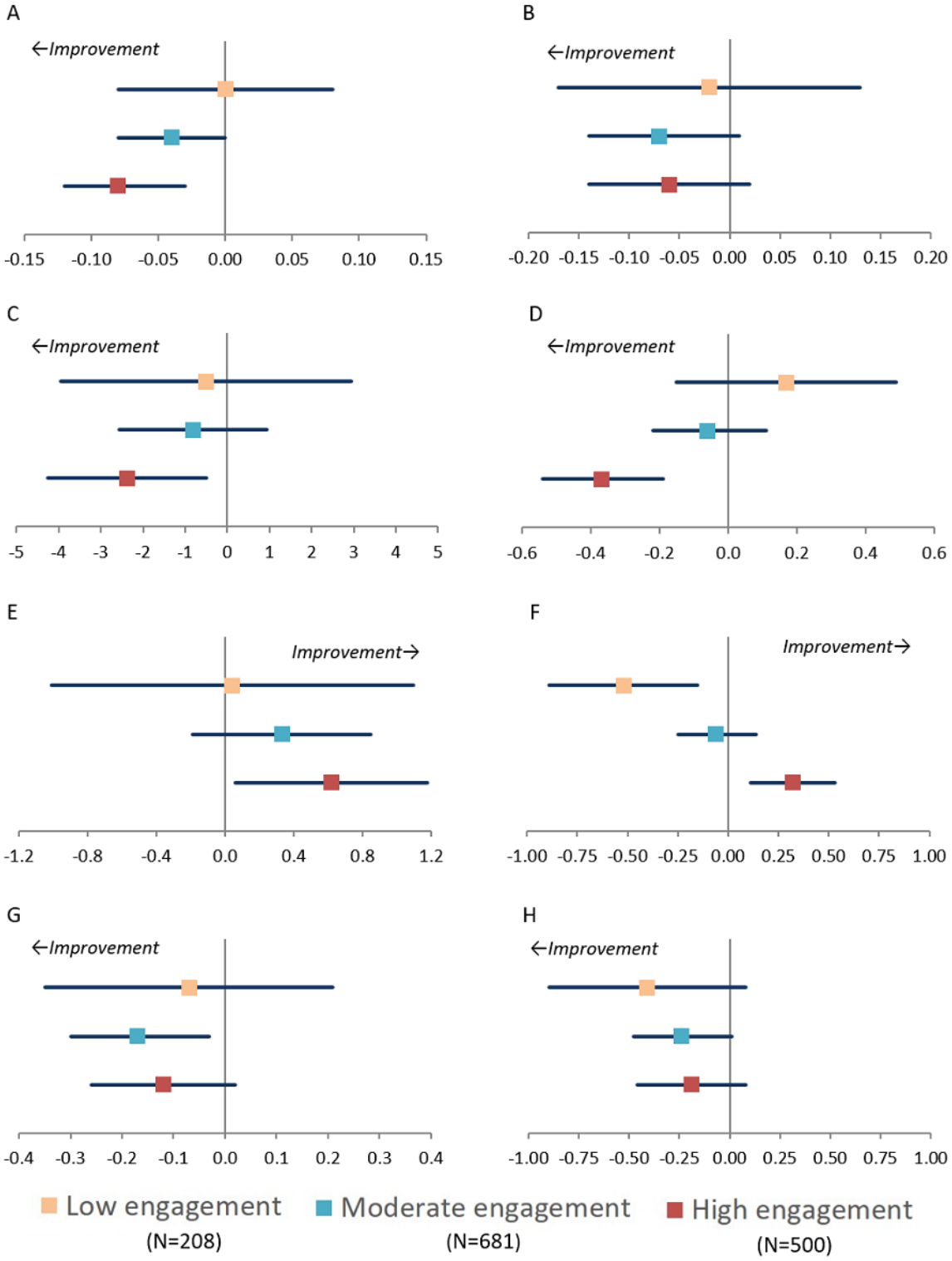
[f]The first instance of nonuse attrition from month 3 onward. The analysis included individuals who had not already undergone an episode of nonuse attrition during the first 2 months. The following variables were included in the initial multivariate model but did not remain significantly associated with late nonuse attrition following a backward stepwise selection procedure: education, current smoking, obesity, age, Mini Mental Status Examination score, verbal fluency score, Stroop score, Rey Auditory Verbal Learning Test recall score, Short Physical Performance Battery score, Partners in Health score, and Mediterranean Diet Adherence Screener nutrition score.

Association Between Engagement and Intervention Outcomes

There was a significantly greater improvement in the HATICE primary outcome measure, comprising systolic blood pressure, BMI, and low-density lipoprotein cholesterol, over 18 months in the high engagement category than in the control group (adjusted mean difference −0.08, 95% CI −0.12 to −0.03; *P*=.001), with an indication of a dose-response effect (Figure 3 and Table S6 in Multimedia Appendix 1; overall *P* value across the 3 adherence groups=.005). Similarly, compared with those in the control group, there was a significantly greater decrease in systolic blood pressure and BMI and significantly less decline in physical activity and Mediterranean Diet Adherence Screener score (all indicating improvement of cardiovascular or dementia risk) in the high engagement category over 18 months (Figure 3 and Table S6 in Multimedia Appendix 1). The results were also numerically, if not significantly, in favor of greater improvement in the other outcome measures, except for Systematic Coronary Risk Estimation–Older People (SCORE-OP), in the high and moderate engagement groups. The results were comparable when the engagement scores were categorized into tertiles (Table S7 in Multimedia Appendix 1).

XSL•FO
RenderX

Figure 3. Adjusted mean difference in 18-month changes in outcome measures in low, moderate, and high engagement categories in the HATICE intervention group compared with control group: (A) HATICE composite z-score (BMI, LDL, and SBP), (B) LDL cholesterol (mmol/L), (C) SBP (mm Hg), (D) BMI (kg/m2), (E) moderate-intense physical activity (hours/week), (F) MEDAS score (range 0-14 points), (G) CAIDE dementia risk score (range 0-15 points), and (H) SCORE-OP (10-year CVD mortality risk). Point estimates are the mean difference in 18-month change compared with the control group. Bars are 95% CIs. Each model was adjusted for baseline age, sex, education, country, physical function, smoking, plans to make lifestyle changes, computer use, and cognition and for baseline score of the outcome of interest if it differed across engagement groups. The HATICE primary outcome measure was a composite score based on the average of 18-month changes in SBP, LDL cholesterol, and BMI z-scores. CAIDE: Cardiovascular Risk Factors, Aging, and Incidence of Dementia; CVD: cardiovascular disease; HATICE: Healthy Ageing Through Internet Counselling in the Elderly; LDL: low-density lipoprotein; MEDAS: Mediterranean Diet Adherence Screener; SCORE-OP: Systematic Coronary Risk Estimation–Older People; SBP: systolic blood pressure.



XSL•FO
RenderX

Discussion

Principal Findings

In an 18-month randomized trial in older adults, compared with those in the control group, those in the intervention group who engaged most with the eHealth intervention designed to encourage lifestyle changes showed significantly greater improvement in objectively and subjectively measured cardiovascular and dementia risk factors. Those with low engagement showed no difference compared with the control group. Participants who reported that they were already working on improving their lifestyle at baseline, or had short-term plans to do so, were more engaged with the intervention. Those who reported not using a computer in the month before baseline were extremely unlikely to engage, irrespective of their intentions regarding lifestyle change.

Although most intervention group participants engaged with the HATICE platform to some extent (eg, 1238/1389, 89.13% set at least one goal), some intervention components, notably the advice and education sections, were used less frequently than others. Interestingly, lifestyle factors (ie, weight loss and physical activity) were the most frequent targets for goal setting, but participants tended to read more advice and education materials when they set a goal relating to a clinical risk factor (ie, cholesterol, blood pressure, or diabetes), suggesting potential differences in engagement depending on underlying motivations.

Strengths and Limitations

Our study provides comprehensive data concerning the engagement of older adults with a tailored digital health intervention over a relatively long period. To date, this population has received little attention in this field. We used data from a large international randomized controlled trial and used multiple objective measures of engagement combined into a composite indicator and a large range of validated predictor and health outcome variables. Our results can be interpreted alongside the qualitative research conducted within the same trial [19,21,36]. A limitation is that, given the difficulty in defining a suitable *dose* of eHealth interventions and a lack of consensus in the literature on how to measure engagement [10], our engagement indicator was arbitrarily defined after trial completion based on the distributions of the different metrics. However, the definition was chosen before conducting any of the comparative analyses presented here, and it adequately captured differences in engagement. Nonetheless, although some participants demonstrated very high levels of engagement with the eHealth intervention, this subsample was very small. Therefore, we could not specifically study the associations between this very high level of engagement and study outcomes. In addition, our participants, who had at least basic computer literacy (due to study eligibility criteria) and had consented to participate in an eHealth behavior change intervention trial, are not representative of the general older population, in which disparities in engagement would likely be greater. However, the multinational context and use of various recruitment methods have increased the diversity of our population.

Comparison With Previous Work

As in younger populations, engagement declined over time in our trial, with a typical sigmoidal pattern of nonuse attrition indicating a *curiosity plateau* followed by a *rejection phase* [8]. Similarly, although older adults were less likely to participate in a web-based chronic disease self-management intervention trial than were younger adults, those who did participate engaged with the intervention in a manner similar to the younger participants [13]. Sustaining engagement may be vital for the long-term effectiveness of digital health interventions, as efficacy declines over time [6]. Automated reminders, and in particular human support, may increase engagement. Adherence to supervised lifestyle interventions appears greater than that to unsupervised interventions in older populations [37], and qualitative research with the HATICE participants underlined the motivational role of the coaches in this trial [36]. Participants also reported that user-friendliness, notably in terms of the attractiveness of the platform, and technical difficulties (eg, login problems) also influenced their engagement with the platform [36]. Integrating digital health interventions into primary care might enhance sustainability and provide additional motivation to older individuals not yet considering lifestyle changes or with reticence regarding such programs [22,36].

Similar to a Finnish computerized cognitive training intervention for older adults [38], and as mentioned in our qualitative work [36], previous level of computer use was the strongest predictor of engagement with our web-based intervention, even in our more contemporary population, in which basic computer literacy was an inclusion criterion. The notion of computer use may reflect both computer literacy and computer access (or quality of access) and is likely to be an indicator of inequalities in access to the eHealth intervention. Indeed, compared with HATICE participants who reported regular use at baseline (N=1344), those who did not regularly use a computer (N=43) were significantly more likely to be older and women and to have a lower level of cognition and education (data not shown). Moreover, Dutch participants were less likely to be regular computer users at baseline and engaged less with the platform than those in France and Finland. This could reflect cultural differences in attitudes toward behavior change, prevention, and research participation [19,21]. Furthermore, self-selection due to recruitment methods may have led the French and Finnish samples to be biased toward more motivated and health-focused individuals. Dutch participants, who were recruited via their general practitioners and were likely influenced by medical authority [19], appeared to be more representative of the general population, notably in terms of education level [17]. Participants' intentions to make lifestyle changes were also associated with platform engagement, and better self-management of chronic conditions was associated with a lower risk of early nonuse attrition. Both relate to self-efficacy, an established predictor of adherence to lifestyle interventions [39].

Higher engagement with the HATICE intervention platform was associated with more favorable changes in the trial's main cardiovascular and dementia risk outcome measures, with evidence of a dose-response effect. Similar results have been reported for eHealth interventions targeting various health

XSL•FO
RenderX

conditions or behaviors in younger populations [10], but the mechanisms (or *mediators*, eg, increased knowledge, motivation, self-efficacy, or affect management) underlying such relationships are not well understood [24,25]. Furthermore, we assessed engagement using an aggregate indicator, suggested to be the most useful measure of engagement [10], but it is important to understand whether measures of frequency (eg, logins) and intensity (eg, number of messages sent or amount of advice and education read) of engagement and passive (eg, reading advice and education) and active (eg, entering measurements and sending messages) platform use all influenced outcomes similarly. Not all forms of engagement with eHealth interventions are necessarily associated with outcomes [29], and the frequency of engagement may be more associated with physical health outcomes [25], whereas intensity is more associated with psychological health outcomes [10]. In addition, it is important to understand whether intervention use is a valid indicator of engagement in behavior change [24] and, if so, through which mechanisms. Finally, engagement with eHealth interventions may encompass more than just objective measures of use, and factors such as interest, enjoyment, and attention may also play a role [24,40].

Conclusions

Engaging older people in an eHealth lifestyle self-management intervention is feasible, and greater engagement is associated with greater improvement in biological and behavioral dementia and cardiovascular risk factors. Further work is required to determine more specifically the strength or type of engagement with such interventions required to obtain a meaningful impact on health outcomes and how best to sustain engagement over time. Our results also suggest disparities in engagement, which, given biases in trial participation, are likely to be accentuated in real-world settings. Older adults with limited computer experience, poorer cognition, and no concrete plans for lifestyle change may require extra support to reach a level of engagement with digital lifestyle interventions that is sufficient to bring about health benefits or require access to alternative methods of intervention delivery to mitigate potential health inequalities that could be associated with the widespread roll-out of digital health interventions in older populations.

Acknowledgments

The research leading to these results received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 779238 and the National Key R&D Programmes of China (2017YFE0118800).

Healthy Ageing Through Internet Counselling in the Elderly is a collaborative project cofunded by the European Union's Seventh Framework Program (2007-2013), under grant agreement 305374.

The members of the Prevention of Dementia using Mobile Phone Applications (PRODEMOS) group are as follows: Edo Richard, Pim van Gool, Eric Moll van Charante, Marieke Hoevenaar-Blom, Esmé Eggink, Melanie Hafdi, and Patrick Witvliet (Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands); Carol Brayne, Linda Barnes, and Rachael Brooks (University of Cambridge, Cambridge, United Kingdom); Wei Wang, Wenzhi Wang, Youxin Wang, and Manshu Song (Capital Medical University, Beijing, China; Edith Cowan University, Perth, Australia); Anders Wimo and Ron Handels (Karolinska Institutet, Stockholm, Sweden); Sandrine Andrieu and Nicola Coley (INSERM UMR1295, Toulouse France); Jean Georges and Cindy Birck (Alzheimer Europe, Luxembourg); Bram van de Groep and Mark van der Meijden (Vital Health Software, Ede, the Netherlands)

The members of the Healthy Ageing Through Internet Counselling in the Elderly group are as follows: Edo Richard, Pim van Gool, Eric Moll van Charante, Cathrien Beishuizen, Susan Jongstra, Tessa van Middelaar, Lennard van Wanrooij, and Marieke Hoevenaar-Blom (Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands); Hilkka Soininen, Tiia Ngandu, and Mariagnese Barbera (University of Eastern Finland, Kuopio, Finland); Miia Kivipelto and Francesca Mangiasche (Karolinska Institutet, Stockholm, Sweden); Sandrine Andrieu, Nicola Coley, and Juliette Guillemont (INSERM–Toulouse University UMR1027, Toulouse, France); Yannick Meiller (Novapten, Paris, France); Bram van de Groep (Vital Health Software, Ede, the Netherlands); and Carol Brayne (University of Cambridge, Cambridge, United Kingdom).

The funders played no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the paper for publication. The researchers were independent from the funders. All authors had full access to the data and take responsibility for the integrity of the data and the accuracy of the data analysis, and all gave approval for publication of the manuscript.

Data Availability

All individual participant data collected during the trial, after deidentification, and study documents, including the study protocol, will be available to academic researchers who provide a methodologically sound proposal, approved by the Healthy Ageing Through Internet Counselling in the Elderly steering group, and agree to collaborate on the analysis with at least one member of the Healthy Ageing Through Internet Counselling in the Elderly study team. Enquiries or proposals should be directed to MHB.

XSL•FO
RenderX

Authors' Contributions

NC, TN, C Beishuizen, MB, MK, HS, WvG, C Brayne, EMVC, ER, and SA made a substantial contribution to the study concept and design. NC, MHB, C Beishuizen, MB, LvW, EMVC, ER, HS, and SA contributed to data collection. MHB, C Beishuizen, and LvW contributed to data management. NC performed the data analysis. NC, LA, MHB, TN, C Beishuizen, MB, LvW, MK, HS, WvG, C Brayne, EMVC, ER, and SA contributed to data interpretation. NC drafted the manuscript, tables, and figures. All authors contributed to the revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figures.
[DOCX File , 95 KB - jmir_v24i5e32006_app1.docx]

References

1. United Nations. World Population Ageing 2017: Highlights. UN iLibrary. 2018 Oct. URL: https://www.un-ilibrary.org/content/books/9789213627457 [accessed 2022-04-27]
2. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
3. Wimo A, Guerchet M, Ali G, Wu Y, Prina AM, Winblad B, et al. The worldwide costs of dementia 2015 and comparisons with 2010. Alzheimers Dement 2017 Jan;13(1):1-7 [FREE Full text] [doi: 10.1016/j.jalz.2016.07.150] [Medline: 27583652]
4. D'Agostino RB, Pencina MJ, Massaro JM, Coady S. Cardiovascular disease risk assessment: insights from Framingham. Glob Heart 2013 Mar;8(1):11-23 [FREE Full text] [doi: 10.1016/j.gheart.2013.01.001] [Medline: 23750335]
5. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Lancet 2020 Aug 08;396(10248):413-446 [FREE Full text] [doi: 10.1016/S0140-6736(20)30367-6] [Medline: 32738937]
6. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
7. Wesselman LM, Hooghiemstra AM, Schoonmade LJ, de Wit MC, van der Flier WM, Sikkes SA. Web-based multidomain lifestyle programs for brain health: comprehensive overview and meta-analysis. JMIR Ment Health 2019 Apr 09;6(4):e12104 [FREE Full text] [doi: 10.2196/12104] [Medline: 30964438]
8. 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] [Medline: 15829473]
9. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
10. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
11. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an Internet intervention for depression. J Affect Disord 2018 Mar 15;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]
12. Mclaughlin M, Delaney T, Hall A, Byaruhanga J, Mackie P, Grady A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180 [FREE Full text] [doi: 10.2196/23180] [Medline: 33605897]
13. Portz JD, LaMendola WF. Participation, retention, and utilization of a Web-based chronic disease self-management intervention among older adults. Telemed J E Health 2019 Feb;25(2):126-131 [FREE Full text] [doi: 10.1089/tmj.2017.0208] [Medline: 29782228]
14. Rueda S, Artazcoz L, Navarro V. Health inequalities among the elderly in western Europe. J Epidemiol Community Health 2008 Jun;62(6):492-498. [doi: 10.1136/jech.2006.059279] [Medline: 18477747]
15. Davidson S. Digital Inclusion Evidence Review 2018. Age UK. 2018 Nov. URL: https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/age_uk_digital_inclusion_evidence_review_2018.pdf [accessed 2022-04-27]
16. Richard E, Jongstra S, Soininen H, Brayne C, Moll van Charante EP, Meiller Y, et al. Healthy Ageing Through Internet Counselling in the Elderly: the HATICE randomised controlled trial for the prevention of cardiovascular disease and cognitive impairment. BMJ Open 2016 Jun 10;6(6):e010806 [FREE Full text] [doi: 10.1136/bmjopen-2015-010806] [Medline: 27288376]

XSL•FO
RenderX

17. Richard E, Moll van Charante EP, Hoevenaar-Blom MP, Coley N, Barbera M, van der Groep A, et al. Healthy ageing through internet counselling in the elderly (HATICE): a multinational, randomised controlled trial. Lancet Digit Health 2019 Dec;1(8):e424-e434 [FREE Full text] [doi: 10.1016/S2589-7500(19)30153-0] [Medline: 33323224]
18. Jongstra S, Beishuizen C, Andrieu S, Barbera M, van Dorp M, van de Groep B, et al. Development and validation of an interactive internet platform for older people: the healthy ageing through internet counselling in the elderly study. Telemed J E Health 2017 Feb;23(2):96-104. [doi: 10.1089/tmj.2016.0066] [Medline: 27463120]
19. Coley N, Rosenberg A, van Middelaar T, Soulier A, Barbera M, Guillemont J, MIND-AD, HATICE groups. Older adults' reasons for participating in an eHealth prevention trial: a cross-country, mixed-methods comparison. J Am Med Dir Assoc 2019 Jul;20(7):843-9.e5. [doi: 10.1016/j.jamda.2018.10.019] [Medline: 30541689]
20. Barbera M, Mangialasche F, Jongstra S, Guillemont J, Ngandu T, Beishuizen C, HATICE study group. Designing an Internet-based multidomain intervention for the prevention of cardiovascular disease and cognitive impairment in older adults: the HATICE trial. J Alzheimers Dis 2018;62(2):649-663. [doi: 10.3233/JAD-170858] [Medline: 29480185]
21. Akenine U, Barbera M, Beishuizen CR, Fallah Pour M, Guillemont J, Rosenberg A, HATICE study group. Attitudes of at-risk older adults about prevention of cardiovascular disease and dementia using eHealth: a qualitative study in a European context. BMJ Open 2020 Aug 06;10(8):e037050 [FREE Full text] [doi: 10.1136/bmjopen-2020-037050] [Medline: 32764085]
22. Beishuizen CR, Akenine U, Barbera M, Rosenberg A, Fallah Pour M, Richard E, et al. Integrating nurses' experiences with supporting behaviour change for cardiovascular prevention into a self-management internet platform in Finland and the Netherlands: a qualitative study. BMJ Open 2019 Jun 06;9(6):e023480 [FREE Full text] [doi: 10.1136/bmjopen-2018-023480] [Medline: 31175194]
23. Hookham G, Nesbitt K. A systematic review of the definition and measurement of engagement in serious games. In: Proceedings of the Australasian Computer Science Week Multiconference. 2019 Presented at: ACSW '19; January 29-31, 2019; Sydney, Australia p. 1-10. [doi: 10.1145/3290688.3290747]
24. 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] [Medline: 27966189]
25. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. J Med Internet Res 2018 Nov 16;20(11):e292 [FREE Full text] [doi: 10.2196/jmir.9397] [Medline: 30446482]
26. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
27. Antypas K, Wangberg SC. An Internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014 Mar 11;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
28. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a Web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172 [FREE Full text] [doi: 10.2196/jmir.2258] [Medline: 23963284]
29. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
30. Power JM, Phelan S, Hatley K, Brannen A, Muñoz-Christian K, Legato M, et al. Engagement and weight loss in a web and mobile program for low-income postpartum women: fit moms/ Mamás Activas. Health Educ Behav 2019 Dec;46(2_suppl):114-123. [doi: 10.1177/1090198119873915] [Medline: 31742447]
31. Stewart AL, Mills KM, King AC, Haskell WL, Gillis D, Ritter PL. CHAMPS physical activity questionnaire for older adults: outcomes for interventions. Med Sci Sports Exerc 2001 Jul;33(7):1126-1141. [doi: 10.1097/00005768-200107000-00010] [Medline: 11445760]
32. Schröder H, Fitó M, Estruch R, Martínez-González MA, Corella D, Salas-Salvadó J, et al. A short screener is valid for assessing Mediterranean diet adherence among older Spanish men and women. J Nutr 2011 Jun;141(6):1140-1145. [doi: 10.3945/jn.110.135566] [Medline: 21508208]
33. Cooney MT, Selmer R, Lindman A, Tverdal A, Menotti A, Thomsen T, SCORE and CONOR investigators. Cardiovascular risk estimation in older persons: SCORE O.P. Eur J Prev Cardiol 2016 Jul;23(10):1093-1103. [doi: 10.1177/2047487315588390] [Medline: 26040999]
34. Kivipelto M, Ngandu T, Laatikainen T, Winblad B, Soininen H, Tuomilehto J. Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study. Lancet Neurol 2006 Sep;5(9):735-741. [doi: 10.1016/S1474-4422(06)70537-3] [Medline: 16914401]
35. Williams R. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. Stata J 2006 Feb 01;6(1):58-82. [doi: 10.1177/1536867x0600600104]
36. van Middelaar T, Beishuizen CR, Guillemont J, Barbera M, Richard E, Moll van Charante EP, HATICE consortium. Engaging older people in an Internet platform for cardiovascular risk self-management: a qualitative study among Dutch

XSL•FO
RenderX

HATICE participants. BMJ Open 2018 Jan 21;8(1):e019683 [FREE Full text] [doi: 10.1136/bmjopen-2017-019683] [Medline: 29358447]
37. Coley N, Ngandu T, Lehtisalo J, Soininen H, Vellas B, Richard E, HATICE, FINGER, and MAPT/DSA groups. Adherence to multidomain interventions for dementia prevention: data from the FINGER and MAPT trials. Alzheimers Dement 2019 Jun;15(6):729-741. [doi: 10.1016/j.jalz.2019.03.005] [Medline: 31047857]
38. Turunen M, Hokkanen L, Bäckman L, Stigsdotter-Neely A, Hänninen T, Paajanen T, et al. Computer-based cognitive training for older adults: determinants of adherence. PLoS One 2019 Jul 10;14(7):e0219541 [FREE Full text] [doi: 10.1371/journal.pone.0219541] [Medline: 31291337]
39. Collado-Mateo D, Lavín-Pérez AM, Peñacoba C, Del Coso J, Leyton-Román M, Luque-Casado A, et al. Key factors associated with adherence to physical exercise in patients with chronic diseases and older adults: an umbrella review. Int J Environ Res Public Health 2021 Feb 19;18(4):2023 [FREE Full text] [doi: 10.3390/ijerph18042023] [Medline: 33669679]
40. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]

Abbreviations

aOR: adjusted odds ratio
CVD: cardiovascular disease
HATICE: Healthy Ageing Through Internet Counselling in the Elderly

Edited by A Mavragani; submitted 13.07.21; peer-reviewed by G Ludden, W LaMendola, K Champion; comments to author 11.10.21; revised version received 08.11.21; accepted 13.12.21; published 09.05.22.

Please cite as:
Coley N, Andre L, Hoevenaar-Blom MP, Ngandu T, Beishuizen C, Barbera M, van Wanrooij L, Kivipelto M, Soininen H, van Gool W, Brayne C, Moll van Charante E, Richard E, Andrieu S, HATICE study group, PRODEMOS study group
Factors Predicting Engagement of Older Adults With a Coach-Supported eHealth Intervention Promoting Lifestyle Change and Associations Between Engagement and Changes in Cardiovascular and Dementia Risk: Secondary Analysis of an 18-Month Multinational Randomized Controlled Trial
J Med Internet Res 2022;24(5):e32006
URL: https://www.jmir.org/2022/5/e32006
doi:10.2196/32006
PMID:35385395

©Nicola Coley, Laurine Andre, Marieke P Hoevenaar-Blom, Tiia Ngandu, Cathrien Beishuizen, Mariagnese Barbera, Lennard van Wanrooij, Miia Kivipelto, Hilkka Soininen, Willem van Gool, Carol Brayne, Eric Moll van Charante, Edo Richard, Sandrine Andrieu, HATICE study group, PRODEMOS study group. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 09.05.2022. 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.

XSL•FO
RenderX

Original Paper

Readability of English, German, and Russian Disease-Related Wikipedia Pages: Automated Computational Analysis

Jelizaveta Gordejeva[1], BSc; Richard Zowalla[1,2,3], MSc; Monika Pobiruchin[2,4], Dr sc hum; Martin Wiesner[1,2], Dipl Inform Med

[1]Department of Medical Informatics, Heilbronn University, Heilbronn, Germany

[2]Consumer Health Informatics SIG, German Association for Medical Informatics, Biometry & Epidemiology (GMDS e. V.), Cologne, Germany

[3]Center for Machine Learning, Heilbronn University, Heilbronn, Germany

[4]GECKO Institute for Medicine, Informatics & Economics, Heilbronn University, Heilbronn, Germany

Corresponding Author:
Martin Wiesner, Dipl Inform Med
Department of Medical Informatics
Heilbronn University
Max-Planck-Str 39
Heilbronn, 74081
Germany
Phone: 49 7131504 ext 6947
Email: martin.wiesner@hs-heilbronn.de

Abstract

Background: Wikipedia is a popular encyclopedia for health- and disease-related information in which patients seek advice and guidance on the web. Yet, Wikipedia articles can be unsuitable as patient education materials, as investigated in previous studies that analyzed specific diseases or medical topics with a comparatively small sample size. Currently, no data are available on the average readability levels of all disease-related Wikipedia pages for the different localizations of this particular encyclopedia.

Objective: This study aimed to analyze disease-related Wikipedia pages written in English, German, and Russian using well-established readability metrics for each language.

Methods: Wikipedia database snapshots and Wikidata metadata were chosen as resources for data collection. Disease-related articles were retrieved separately for English, German, and Russian starting with the main concept of *Human Diseases and Disorders* (German: *Krankheit*; Russian: Заболевания человека). In the case of existence, the corresponding International Classification of Diseases, Tenth Revision (ICD-10), codes were retrieved for each article. Next, the raw texts were extracted and readability metrics were computed.

Results: The number of articles included in this study for English, German, and Russian Wikipedia was n=6127, n=6024, and n=3314, respectively. Most disease-related articles had a Flesch Reading Ease (FRE) score <50.00, signaling difficult or very difficult educational material (English: 5937/6125, 96.93%; German: 6004/6022, 99.7%; Russian: 2647/3313, 79.9%). In total, 70% (7/10) of the analyzed articles could be assigned an ICD-10 code with certainty (English: 4235/6127, 69.12%; German: 4625/6024, 76.78%; Russian: 2316/3314, 69.89%). For articles with ICD-10 codes, the mean FRE scores were 28.69 (SD 11.00), 20.33 (SD 9.98), and 38.54 (SD 13.51) for English, German, and Russian, respectively. A total of 9 English ICD-10 chapters (11 German and 10 Russian) showed significant differences: chapter F (FRE 23.88, SD 9.95; P<.001), chapter E (FRE 25.14, SD 9.88; P<.001), chapter H (FRE 30.04, SD 10.57; P=.049), chapter I (FRE 30.05, SD 9.07; P=.04), chapter M (FRE 31.17, 11.94; P<.001), chapter T (FRE 32.06, SD 10.51; P=.001), chapter A (FRE 32.63, SD 9.25; P<.001), chapter B (FRE 33.24, SD 9.07; P<.001), and chapter S (FRE 39.02, SD 8.22; P<.001).

Conclusions: Disease-related English, German, and Russian Wikipedia articles cannot be recommended as patient education materials because a major fraction is difficult or very difficult to read. The authors of Wikipedia pages should carefully revise existing text materials for readers with a specific interest in a disease or its associated symptoms. Special attention should be given to articles on *mental, behavioral,* and *neurodevelopmental disorders* (ICD-10 chapter F) because these articles were most difficult to read in comparison with other ICD-10 chapters. Wikipedia readers should be supported by editors providing a short and easy-to-read summary for each article.

(J Med Internet Res 2022;24(5):e36835) doi:10.2196/36835

XSL•FO
RenderX

KEYWORDS

readability; health literacy; health education; Wikipedia

Introduction

Overview

Many people consult the internet as a rapidly accessible resource to find information [1,2]. This applies to patients who want to educate themselves about a disease in a personal or family context [3-5]. Studies have shown that the internet has become a popular source of information for patients [6,7].

According to the available rankings, the web-based encyclopedia Wikipedia is a popular domain worldwide [8,9]. The web-based encyclopedia also appears among the top websites when searching for health-related information on search engines such as Google [10,11]. In December 2021, the English version of Wikipedia contained 6,423,416 articles [12]. It is a popular knowledge base that is consulted by many users to find out more about diseases and conditions, as well as for self-education purposes [13,14]. The target groups of Wikipedia are heterogeneous and include patients, students, practitioners, and the public [15].

Wikipedia articles can be written and edited by everyone, which adds to their popularity [15]. Farič et al [16] found that health-related content on Wikipedia is created by both health specialists and laypeople. Wikipedians are driven by values and beliefs, intrinsic motivation, and a certain sense of obligation [16]. Although Wikipedia is a popular resource for accessing medical knowledge [10,17], its readability is not assessed, quality assured, or controlled before publishing. This can lead to articles being difficult to read and understand [10,18,19], which can result in a lack of comprehensibility and an inability to help patients bridge the health literacy gap [20,21].

Patient education, for example, through internet searches, is an important step in medical compliance and patient empowerment [20,22,23]. Thus, it can influence the health care process and patient-physician relationship either positively or negatively [24,25]. However, to use texts as patient education materials, they must be understandable and easy to read. Readability can be defined as the number of school grades or years of formal education a person has received. In the United States, the recommended grade level for patient education materials is 7 to 8 [26].

The readability of a text can be computed using several established metrics. Several formulas are available for English texts [27-31], with adaptions for the German [32] or Russian language [33]. However, manually calculating the readability of texts is a resource- and time-consuming task.

This study assessed the readability of disease-related Wikipedia articles written in 3 different languages. Using an automated computation approach, one of our aims was to assess whether Wikipedia articles are suitable as patient education materials.

Related Work

Several studies have analyzed the readability of health-related Wikipedia pages. Before this study, we conducted a systematic literature review to assess how readability metrics have been used to evaluate health-related Wikipedia articles. The details of the review are included in Multimedia Appendix 1 [10,11,17-19,34-46]. In total, 31 articles were closely evaluated. In most publications, the readability of texts was assessed using web-based accessible software (eg, [47]), to which texts of English Wikipedia articles were manually copied. Furthermore, only articles on certain diseases or health-related topics were analyzed (eg, anatomy [34] or pediatric ophthalmology [18] articles). In general, Wikipedia pages were difficult to read. Some selected publications related to this study are presented in the following paragraphs.

Brigo et al [35] assessed the readability of 41 Wikipedia articles on epilepsy. The selected Wikipedia articles were divided into two categories: (1) articles related to epilepsy (n=23) and (2) articles related to antiepileptic drugs (n=18). The authors found that average Flesch Reading Ease (FRE) values for these articles were 30.2 (SD 8.1) and 19.6 (SD 7.6) for epilepsy and antiepileptic drugs, respectively. Both values corresponded to texts that were difficult to read and understand. Other classic readability metrics were also calculated. On average, all analyzed Wikipedia articles "[...] correspond to a 14th academic grade level (14.3±1.7) and to 16.4±2.0 years of formal education required to easily understand the text on the first reading."

In 2020, Suwannakhan et al [34] selected 40 *anatomy* articles from Wikipedia to analyze their readability. The assessment showed that, on average, these articles were difficult to understand and required at least a college education level (FRE: mean 42.4, SD 10.8; Flesch-Kincaid Grade Level [FKGL]: mean 12.3, SD 2.1).

In a recent study, Handler et al [36] compared the readability of Wikipedia articles on *pelvic floor disorders* with that of patient education leaflets. The authors collected 30 Wikipedia articles and 29 leaflets. They found that Wikipedia articles (Simple Measure of Gobbledygook [SMOG] 12.0) were significantly (P<.001) harder to read than the patient education material (SMOG 3.4). The authors also reported readability values for Wikipedia articles in different categories. The collected articles corresponded to a college-level education needed to adequately comprehend the text.

Hutchinson et al [10] investigated the readability of texts available on the web regarding *internal medical* diagnoses. In this study, Google was used to collect data. Wikipedia appeared among the top 5 resulting websites. The authors stated that texts acquired from Wikipedia had an average grade level of 14.6, which was the highest value among all sources.

Similarly, John et al [18] compared the readability of different information available on the web on *pediatric ophthalmology*. In this context, a Google search was performed, and Wikipedia, among other websites, was searched internally for relevant articles on this topic. A total of 34 articles were retrieved, including 10 (29%) from Wikipedia. The authors also found that Wikipedia was the most difficult to understand in

XSL•FO
RenderX

comparison with other analyzed resources, with an average grade level of 17.4 (SD 1.18).

In 2020, Shetty et al [19] conducted a search for patient education material available on the web regarding *otitis media*. The authors then assessed the readability of 6 selected websites, including Wikipedia, with 24 patient education pages. Across all investigated resources, Wikipedia had the highest reading level (Gunning Frequency of Gobbledygook [FOG] 15.95, SMOG 14.6, FKGL 12.5, Coleman-Liau Index [CLI] 12.64, and Automated Readability Index [ARI] 11.92).

Of the 31 studies, 8 (26%) also compared the readability of Wikipedia with other available sources (eg, patient education brochures and websites) [11,37-43]. All of them reported Wikipedia to be the most difficult to read among the assessed sources.

In 2014, Kräenbring et al [44] analyzed the readability of Wikipedia articles on pharmacology written in German. In total, 100 curricular drugs were selected from the relevant textbooks. In the context of this study, only Wikipedia articles that overlapped with the information provided in the textbooks were included (n=95). The authors reported the modified FRE for the German language by Amstad [32] and the Vienna Formula (*Wiener SachTextFormel* [WSTF]) readability metrics for Wikipedia and textbooks, respectively: "[...]R_{Amstad}: 7.1±1.7 vs. 7.4±1.8, P=0.9; $R_{1.\ WSTF}$: 15.4±0.5 vs.14.5±0.2, P=0.07." In their study, they found no significant difference between the readability of Wikipedia and the selected textbooks. Both sources provided information that was difficult to read and required tertiary education for an adequate understanding of the material.

Aims of the Study

Compared with previous research in this field, this study does not focus on 1 particular medical subfield but includes all disease-related Wikipedia pages in 3 languages: English, German, and Russian [48-50]. The authors decided to focus on the Wikipedia category *Human Diseases and Disorders* (German: *Krankheit*; Russian: Заболевания человека) because all articles related to diseases are associated with this category.

In this context, four aims were defined: (1) to automatically collect articles from Wikipedia related to the category *Human Diseases and Disorders*, acquire the current state, and report descriptive statistics such as the number of articles, sentences, and words; (2) to categorize them into distinguishable medical subfields using the International Classification of Diseases, Tenth Revision (ICD-10) [51]; (3) to automatically evaluate the readability of retrieved articles with established readability metrics per language; and (4) to assess and compare the text difficulty among medical fields and languages.

In the context of the second aim, ICD-10 was chosen because (1) it has been widely adopted in many health care systems worldwide ("all WHO member states" [52]), and (2) respective codes are provided in many disease-related Wikipedia articles [53-55] to provide a precise reference to a stable classification system.

Methods

Study Design

This study comprised 2 stages. To answer the first 2 research aims, the authors separately collected articles from English, German, and Russian Wikipedia that belong to the category *Human Diseases and Disorders* (German: *Krankheit*; Russian: Заболевания человека). Furthermore, the data collection for each language was expanded based on the articles retrieved from the other 2 Wikipedia domains. Next, ICD-10 codes were retrieved automatically from Wikidata, a central, structured data knowledge base of related resources (ie, Wikipedia), which can be read by machines. For each language, the included articles were separated into two groups: (1) articles that were retrieved and had an ICD-10 code assigned (group A) and (2) articles without an ICD-10 code (group B).

Starting in 2020, a transition toward International Classification of Diseases, Eleventh Revision (ICD-11), [56] began [57], which is already being disseminated in the English Wikipedia edition; however, German or Russian Wikipedia authors still refer to ICD-10 in 2021. With no ICD-11 codes available for all 3 languages, the ICD-10 was chosen for stable comparisons.

Articles included in group A referred to *Human Diseases and Disorders* for which a corresponding ICD-10 code was either directly annotated or could be resolved. However, articles in group B were not excluded for several reasons. Some articles in this group referred to human diseases; however, the ICD-10 was not referred to by any of the authors of the article. Some articles were associated with a disease (eg, symptoms, root causes, and physiological processes) and did not have a specific ICD-10 code. However, those were nevertheless of importance to readers; for example, *cough* (German: *Husten*; Russian: Кашель) and *cytokine storm* (German: *Zytokinsturm*; Russian: Цитокиновый шторм).

In the subsequent stage, the authors used collected data from the first stage to perform an automated readability analysis to answer research aims 3 and 4.

Study Setting

A total of 3 static snapshots of Wikipedia's database, dated June 30, 2021, were used to build a category graph for *Human Diseases and Disorders* per language. Each of the 3 language-specific databases was queried to retrieve and analyze the readability of all relevant plain texts. For a detailed description of the preprocessing steps, see the *Computational Processing Steps* section.

Readability Analysis

Definition

Readability [58] refers to the properties of written text with respect to readers' competence, motivation, and understanding of a document [59]. It reflects the (1) complexity of a text's structure, (2) sentence structure, and (3) chosen vocabulary.

FRE Scale

A well-known readability formula for English is the FRE metric [27]. To compute the FRE metric for a given text, the average

XSL•FO
RenderX

sentence length (ASL) and average syllables per word (ASW) must be calculated. FRE relies on the observation that short words or sentences are usually easier to understand than longer words.

For this analysis, three versions of FRE were applied: (1) the original metric developed for the English language [27], (2) the modified FRE for the German language developed by Amstad [32], and (3) the modified FRE for the Russian language developed by Solovyev et al [33], as shown in the following equations:

$$FRE = 206.835 – (1.015 \times ASL) – (84.6 \times ASW)$$

$$FRE = 180 – ASL – (58.5 \times ASW)$$

$$FRE = 208.7 – (2.6 \times ASL) – (39.2 \times ASW)$$

FKGL Metric

Another widely used readability metric for the English language is the FKGL readability test [60]. It is a modified version of the FRE and was developed to assess readability on the scale of US school grades. This formula, similar to FRE, is based on ASW and ASL.

In this study, the authors used two versions of FKGL: (1) the original FKGL metric for the English language and (2) the modified FKGL for the Russian language developed by Solovyev et al [33], as shown in the following equations:

$$FKGL = (0.39 \times ASL) + (11.8 \times ASW) – 15.59$$

$$FKGL = (0.36 \times ASL) + (5.76 \times ASW) – 11.97$$

The Gunning FOG Index

Gunning FOG is a measure of readability that also relies on the fact that shorter words and sentences are easier to understand. It was developed by Gunning [28] to measure the readability of English text. The formula is based on ASL and the percentage of hard words—that is, ≥3 syllables—in the text, as shown in the following equation:

$$Gunning\ FOG = 0.4 \times (ASL + percentage\ of\ hard\ words)$$

SMOG Grade Level

Another established readability formula for the English language is SMOG. It was derived by McLaughlin [29]. It is based on the count of polysyllabic words (p) (ie, ≥3 syllables) in samples of 30 sentences, as shown in the following equation:



The ARI Metric

ARI is a readability scale derived from sentence difficulty and word difficulty. It was proposed by Senter and Smith [61]. Unlike the aforementioned metrics, word difficulty is calculated based on the character count of the word, not the syllable count, as shown in the following equation:

$$ARI = 4.71 \times (characters / words) + 0.5 \times (words / sentences) – 21.43$$

The CLI Metric

In the Coleman-Liau readability formula [30], similar to the ARI, the difficulty of a word is calculated based on the average number of letters per 100 words (L). In contrast, sentence difficulty is derived from the average number of sentences per 100 words (S), as shown in the following equation:

$$CLI = 0.0588 \times L – 0.296 \times S – 15.8$$

Vienna Formula (WSTF)

The authors applied this metric to measure the readability of German texts. In contrast to the FRE, the Vienna formula (WSTF) is not an adapted version of the German language. Instead, it relies on the work of Bamberger and Vanacek [62], who conducted an analysis based on German texts. They derived at least five versions of the Vienna formula for prose and nonfictional texts. Typically, the fourth WSTF was used for text analysis. This metric also relies on ASL and the proportion of (complex) words with ≥3 syllables (MS), as shown in the following equation:

$$Fourth\ WSTF = 0.2656 \times ASL + 0.2744 \times MS – 1.6939$$

Difficulty

Most metrics, apart from FRE, output school grades. This indicates the degree of education required to understand the text. For instance, a grade of 10 corresponds to an easily readable text, which is suitable for readers educated to at least 10th grade and corresponds to the age of 15 to 16 years in the US school system. The higher the grade level, the more difficult it is to understand.

The FRE metric yields values on a scale of 0 to 100; lower values indicate a text with a low level of readability that is difficult to read, whereas higher values reflect an easily readable text.

Computational Processing Steps

Stage 1: Data Collection

The following subsections describe the steps that were conducted in stage 1 to build a data collection of relevant disease-related articles to be included in the study.

Step 1: Graph-Based Data Retrieval From Wikipedia

We used a static snapshot of the Wikipedia database for the 3 languages of interest obtained using the website [63]. Next, we constructed a graph data structure for each localization starting at the main concept of *Human Diseases and Disorders* (German: *Krankheit*, Russian: Заболевания человека). Graph statistics for each language (ie, the number of nodes and edges) can be found in Multimedia Appendix 2.

This graph contains typed *nodes* that correspond to the structure of Wikipedia and how different articles are interlinked among each other, which is referred to as edges. Subsequently, to filter articles that are related to the concept of interest according to Wikipedia's categorization but are irrelevant for this study, a 4-fold filtering pipeline was used:

XSL•FO
RenderX

1. A wildcard-based category name filter to exclude full subcategories (eg, an overview list of *diseases and disorders by country* or *people with rare diseases*); see Multimedia Appendix 3
2. A given name filter to exclude articles about persons.
3. A geographical filter for articles related to specific countries, cities, or locations.
4. A stop words filter to target all articles that were not excluded by previous filters but were nevertheless not of interest for further evaluation (eg, different disease-related organizations and international disease days).

For every remaining article, the attributes *title*, *Wikipedia page ID*, and *text content* were collected for subsequent analyses.

Step 2: Article Retrieval From Wikidata

Article titles originating from English Wikipedia collected in *step 1* were used to retrieve corresponding articles from the other 2 Wikipedia editions (German and Russian). For this step, the authors used an open-source Java library, *Wikidata Toolkit* (version 0.12.1; Wikimedia) [64]. English titles were used to check for corresponding articles linked to German or Russian Wikipedia pages. If this was the case, the corresponding articles were retrieved and added to the respective language for data collection. Analogously, this process was conducted for the German and Russian articles collected in the first step.

These steps allowed us to balance the differences between collected articles from 3 different Wikipedia domains because of their different category structures and relations.

Step 3: ICD-10 Code Retrieval From Wikidata

The Wikidata resource comprises a structured set of metadata that can be found in related resources; for example, Wikipedia. This allowed the automated retrieval of corresponding ICD-10 codes for those articles for which the original authors or editors did not annotate an ICD-10 code for the Wikipedia page for a certain disease. This is provided through the *P4229* property in the Wikidata knowledge base.

Given every article from step 2, the processing software automatically checked whether an ICD-10 code was provided, and if so, it was added to the respective article in the study's data collection.

For a subset of articles, the ICD-10 code was provided in merely 1 or 2 of 3 Wikipedia editions. In these cases, the available code was automatically added to the remaining corresponding articles, for which no ICD-10 code was found in the original Wikipedia database snapshot.

Some articles were identified as being associated with multiple ICD-10 codes from different chapters. For example, the article on *air embolism* was assigned ICD-10 codes O88.0 (obstetric air embolism) and T79.0 (air embolism [traumatic]). Therefore, each multi-associated article was allocated to all of its available ICD-10 main chapters. Each duplicate was assigned only to 1 ICD-10 main chapter. For instance, the article on *air embolism* was represented in the data collection twice: in ICD-10 chapters O and T.

At the end of the first stage, a collection of Wikipedia articles related to *Human Diseases and Disorders* with the following data was retrieved for the 3 languages: title, Wikipedia page ID, text content, and ICD-10 code in case available or resolved through Wikidata.

Stage 2: Data Analysis

In the second stage of the study, the collected data were analyzed.

For readability computations, the same analysis framework and related processing steps as presented in the study by Wiesner et al [65] was used.

In the context of this study, the raw texts of Wikipedia articles were used as input. Next, all readability metrics described in the *Readability Analysis* section were computed. A vocabulary analysis was not performed in this study.

The analysis was conducted on a Mac OS 10.15.7 64-bit computer with Java Runtime Environment (version 11.0.11; Oracle Corporation).

Statistical Analysis

Data were analyzed using the statistical analysis software R (version 3.6.3; The R Foundation; February 29, 2020) on a MacOS 10.15.7 64-bit computer. The R package *ggplot2* [66] was used for visualization.

Several test scenarios were identified: (1) testing the readability values of 1 language pairwise against the other 2 languages to investigate whether there are significant differences in readability between languages, (2) testing the readability values of each ICD-10 chapter against the mean of all the articles per language, and (3) testing the readability values of each ICD-10 chapter against the recommended readability level of 7 (in the US grade scheme) for patient education material (only for English articles).

For the first and second scenarios, an unpaired 2-tailed *t* test was performed with the following test hypotheses, as shown in the following equations:

$$H_0: \mu_1=\mu_2$$

$$H_a: \mu_1\neq\mu_2$$

For the third scenario, an unpaired 1-tailed *t* test was performed with the hypothesis that articles from all ICD-10 chapters in English (μ_1) have significantly lower readability and thus a higher grade level than the recommended grade level (μ_2), as shown in the following equations:

$$H_0: \mu_1\leq\mu_2$$

$$H_a: \mu_1>\mu_2$$

For all 3 scenarios, a significance level of α=.05 was chosen. For the first scenario, *P* values were Holm adjusted [67] because multiple *t* tests were conducted with the same sample.

ICD-10 chapters were only included in the comparative analyses if the sample size was >25. This restriction was applied to ensure the requirements of the *t* test.

For statistical tests, the values of FRE were used because this was the only readability metric that could be computed for all

XSL•FO
RenderX

languages with a compatible scale. This allowed for a comparison of text readability of different languages.

Ethics Approval

This study does not include any studies with human participants performed by any of the authors. For this reason, no formal ethics approval was obtained for this study.

Results

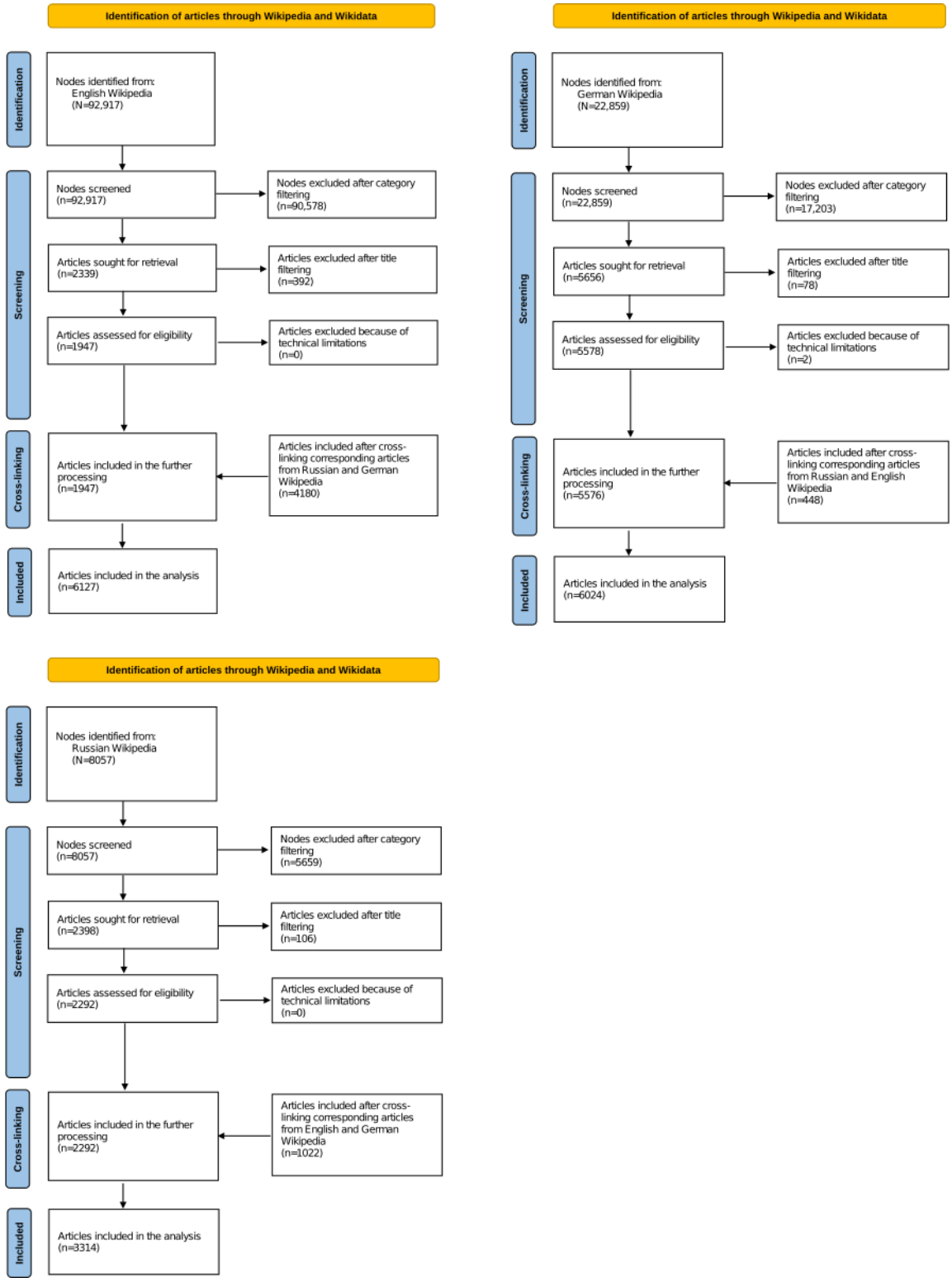
Main Results

Wikipedia article selection and readability analysis were conducted on November 25, 2021 (German), November 30, 2021 (English), and December 3, 2021 (Russian).

After the application of filters, 1947 articles were collected for English, 5576 for German, and 2292 for Russian Wikipedia. These titles were used as input for *step 2* of the data collection (see the *Methods* section). In total, the number of articles included for further readability analysis for English, German, and Russian Wikipedia was n=6127, n=6024, and n=3314, respectively. A detailed summary per language is depicted in Figure 1.

XSL•FO
RenderX

Figure 1. Flow diagrams for the process of data collection for English, German, and Russian.



Sample Characteristics

The retrieved articles were categorized based on their ICD-10 codes into two groups: A and B.

The distribution of articles for each language in these groups is presented in Table 1.

On average, articles from English Wikipedia were the longest with regard to the number of sentences per article (group A: mean 68.49, SD 72.96; group B: mean 52.28, SD 62.02) and the number of words per article (group A: mean 1465.40, SD 1622.97; group B: mean 1213.17, SD 1497.12). Russian Wikipedia articles had the highest number of complex words per article (group A: mean 514.52, SD 734.08; group B: mean 353.89, SD 483.40).

Detailed statistics on the number of sentences, words, and complex words per language can be found in Table 2.

XSL•FO
RenderX

Table 1. Distribution of collected articles in groups A and B for English, German, and Russian languages.

Language	Total, N	Group A, n (%)	Group B, n (%)
English	6127	4235 (69.12)	1892 (30.88)
German	6024	4625 (76.78)	1399 (23.22)
Russian	3314	2316 (69.89)	998 (30.11)

Table 2. Range and mean (SD) values for the number of sentences, words, and complex words in groups A and B for English, German, and Russian languages.

Number of words and sentences and statistics	English, mean (SD; range)		German, mean (SD; range)		Russian, mean (SD; range)	
	Group A	Group B	Group A	Group B	Group A	Group B
Number of sentences	68.49 (72.96; 1-707)	52.28 (62.02; 1-466)	41.73 (61.56; 1-1032)	32.56 (51.15; 1-653)	49.70 (65.95; 2-826)	33.48 (44.63; 2-662)
Number of words	1465.40 (1622.97; 18-14,858)	1213.17 (1497.12; 15-11,987)	708.55 (1180.88; 16-21,531)	582.74 (1004.05; 18-11,515)	837.31 (1233.99; 31-16,352)	589.92 (815.02; 21-8790)
Number of complex words	362.77 (400.89; 3-3694)	299.16 (377.56; 4-3198)	274.91 (441.03; 9-6895)	216.65 (371.51; 4-4597)	514.52 (734.08; 20-9317)	353.89 (483.40; 15-5412)

Readability Analysis

Overview

Readability analysis was performed for every article collected in stage 1 of this study. However, 5 articles in total (English: n=2, 40%; German: n=2, 40%; and Russian: n=1, 20%) were excluded from the analysis because of technical inability to compute the readability metrics for these articles.

Overall, 693 (English: n=212, 30.6%; German: n=390, 56.3%; and Russian: n=91, 13.1%) articles were identified as being associated with multiple ICD-10 codes from different chapters. Consequently, in the following subsections, the reported number of articles in different ICD-10 chapters and groups can differ from the data reported in Table 1.

The distribution of article difficulty for groups A and B, as well as per ICD-10 chapter, can be found in Multimedia Appendices 4-6. Box plots depicting value differences between groups A and B, as well as among ICD-10 chapters, can be found in Multimedia Appendices 7-9 for each computed readability metric.

English Wikipedia

All included articles from English Wikipedia were analyzed according to the readability metrics FRE, FKGL, SMOG, ARI, CLI, and Gunning FOG.

The highest number of articles was assigned to ICD-10 chapter Q (740/4471, 16.55%). The lowest number of articles was associated with chapter W (4/4471, 0.09%). In the context of a low number of articles per ICD-10 chapter, the respective chapters were excluded from the comparative analysis (see *Statistical Analysis* subsection).

The average number of sentences per article varied from 42.42 (ICD-10 chapter Q, SD 45.45) to 110.43 (ICD-10 chapter B, SD 97.90); the average number of words varied from 888.76 (ICD-10 chapter Q, SD 984.97) to 2346.28 (ICD-10 chapter F, SD 2216.12).

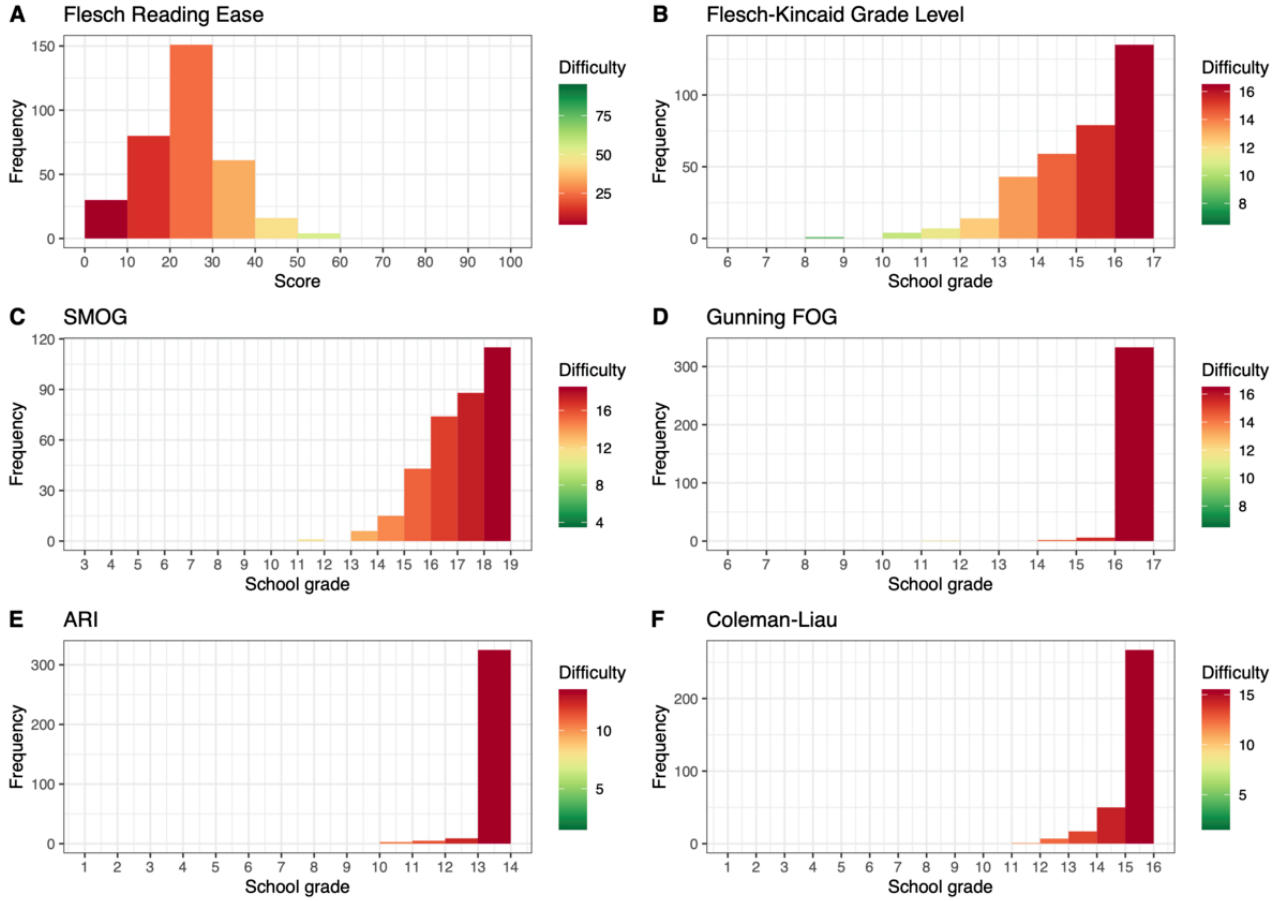
Chapter F had the highest grade level scores (FKGL 15.33, SD 1.53; ARI 13.87, SD 0.51; CLI 15.49, SD 0.85; SMOG 17.22, SD 1.38; and Gunning FOG 16.92. SD 0.42) and the lowest FRE score of 23.88 (SD 9.95) in comparison with other ICD-10 chapters, which indicates difficulty in reading texts.

The articles that were relatively easy to read were in chapter S with an FRE score of 39.02 (SD 8.22), and all grade level indices were <16 (FKGL 12.78, SD 1.58; ARI 13.22, SD 1.03; CLI 14.06, SD 1.15; SMOG 14.66, SD 1.34; and Gunning FOG 15.86, SD 1.29).

Figures 2 and 3 depict the distribution of readability values for each article in chapters F and S, respectively. Each subfigure represents the values computed using different metrics for the same sample. In total, 76.3% (261/342) of the articles in chapter F had an FRE value <30.00; that is, they were very or extremely difficult to read and required a tertiary degree of education for adequate comprehension of the text. In chapter S, only 8% (6/79) of the articles had an FRE score <30.00, whereas 84% (66/79) of the articles had an FRE value between 30.00 and 50.00.

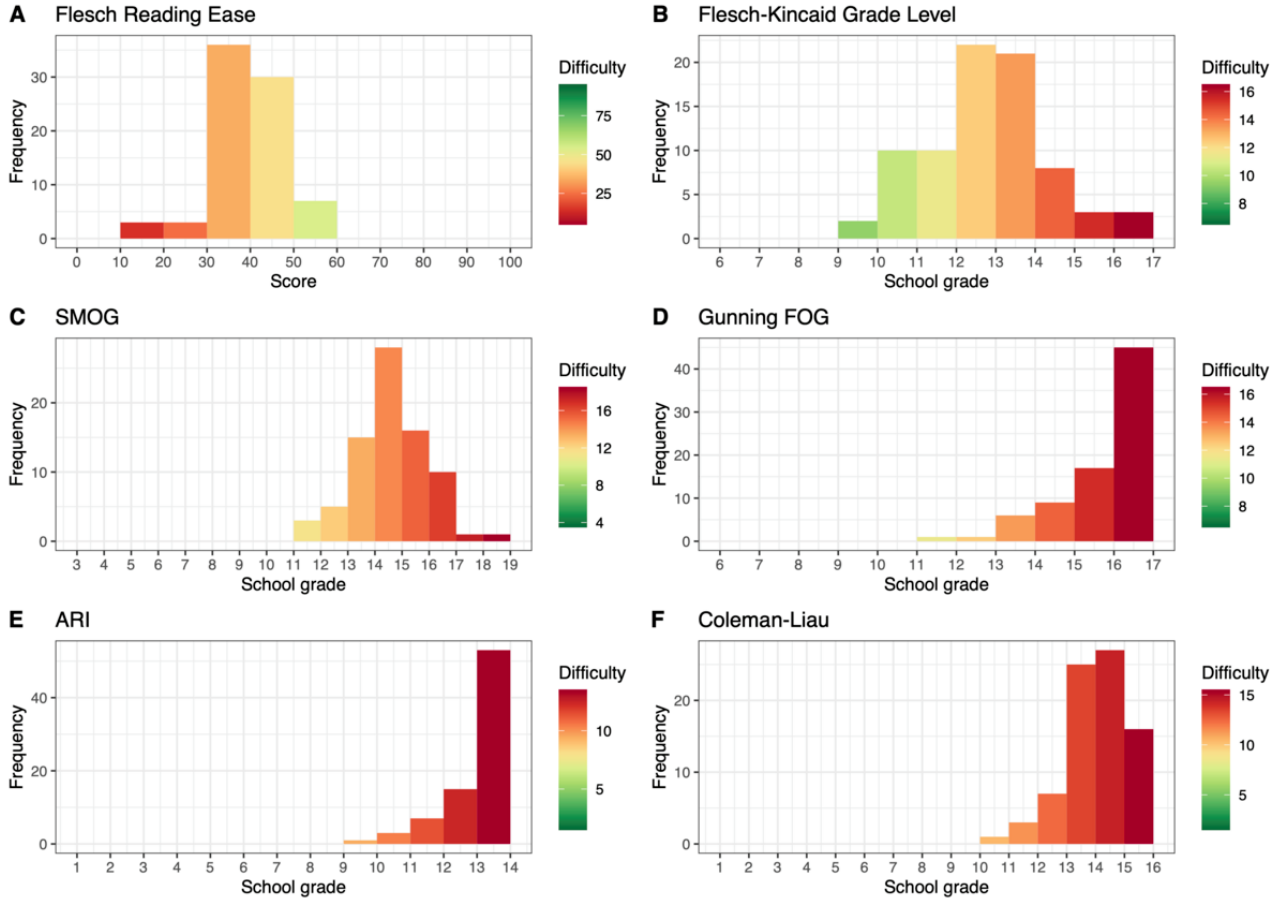
XSL•FO
RenderX

Figure 2. Distribution of values of all computed readability metrics (English) for articles from the International Classification of Diseases, Tenth Revision, chapter F (ICD-F). ARI: Automated Readability Index; FOG: Frequency of Gobbledygook; SMOG: Simple Measure of Gobbledygook.



XSL•FO
RenderX

Figure 3. Distribution of values of all computed readability metrics (English) for articles from the International Classification of Diseases, Tenth Revision, chapter S (ICD-S). ARI: Automated Readability Index; FOG: Frequency of Gobbledygook; SMOG: Simple Measure of Gobbledygook.



German Wikipedia

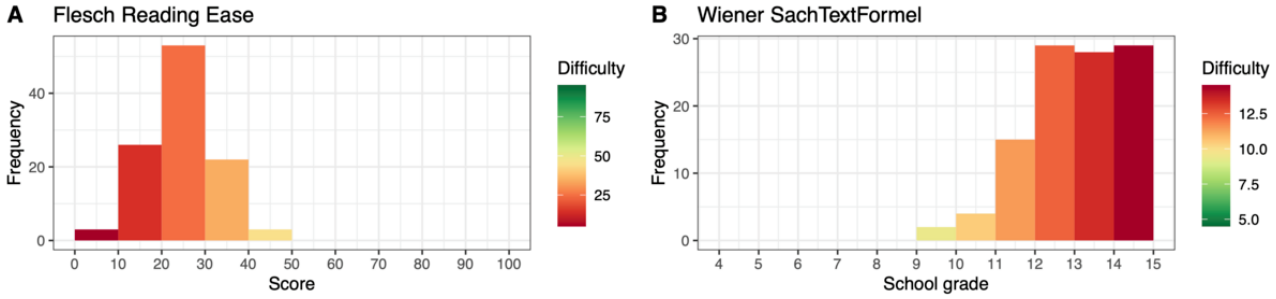
All included articles from German Wikipedia were analyzed according to the readability metrics FRE and WSTF.

The highest number of articles was assigned to ICD chapter Q (1030/5092, 20.23%). The lowest number of articles were in chapter X (3/5092, 0.06%). The lowest average number of sentences per article was in chapter Q (26.28, SD 27.45) and the highest was in chapter B (74.04, SD 94.26). The average word count varied from 398.84 (ICD-10 chapter Q, SD 526.23) to 1337.13 (ICD-10 chapter F, SD 1880.59).

For the FRE metric, the most difficult to read was chapter E, with a score of 17.43 (SD 9.80), and the one that was relatively easy to read was chapter S (24.60, SD 8.22). The highest WSTF value was observed in chapter I (13.62, SD 1.13) and the lowest was in chapter B (13.01, SD 1.24).

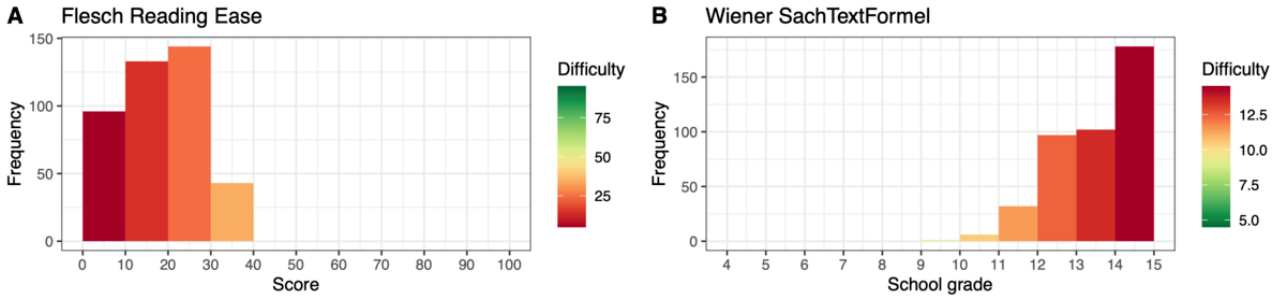
Figures 4 and 5 depict the distributions of the FRE and WSTF values for each article in chapters S and E, respectively. In chapter E, 89.7% (373/416) of the articles were very and extremely difficult to read (FRE<30.00) compared with 76.7% (82/107) of the articles in chapter S, which had an FRE score <30.00.

Figure 4. Distribution of Flesch Reading Ease and Fourth Vienna Formula values for German articles from the International Classification of Diseases, Tenth Revision, chapter S (ICD-S).



XSL•FO
RenderX

Figure 5. Distribution of Flesch Reading Ease and Fourth Vienna Formula values for German articles from the International Classification of Diseases, Tenth Revision, chapter E (ICD-E).



Russian Wikipedia

All included articles from Russian Wikipedia were analyzed according to the readability metrics FRE and FKGL.

The highest number of articles was assigned to ICD chapter F (275/2417, 11.38%); the lowest numbers were found for chapters W and Y (each 1/2417, 0.04%). The average sentence count per article varied from 30.42 (chapter Q, SD 32.05) to 80.60 (chapter A, SD 80.93); the average word count distribution was 489.13 (SD 571.41) to 1289.28 (SD 1468.39) in chapters Q and B, respectively.

Chapter E was, on average, the most difficult to understand (FRE 33.66, SD 12.73, and FKGL 13.35, SD 1.73), whereas the easiest to read articles were in chapter O, with an FRE score of 44.06 (SD 10.73) and an FKGL of 11.88 (SD 1.53), which were the highest average values for these metrics per chapter among all analyzed languages.

Figures 6 and 7 depict the distributions of the FRE and FKGL values for each article in chapters E and O, respectively. In total, 33.4% (80/240) of the articles in chapter E had an FRE value <30.00 (very or extremely difficult to read), whereas 57.5% (138/240) scored on a scale between 30.00 and 50.00 (difficult to read). A value >50.00 was observed for 9.2% (22/240) of the chapter E articles.

In chapter O, most articles (36/41, 8%) were difficult to read and required a college degree for comprehension. In total, 27% (11/41) had an FRE value >50.00, whereas 10% (4/41) had a value <30.00.

The following tables report the mean values and SDs of the readability metrics for each locale. The mean was calculated for groups A and B and every ICD-10 chapter (Tables 3-8). The tables also show the number of articles included in each category.

Figure 6. Distribution of Flesch Reading Ease and Flesch-Kincaid Grade Level values for Russian articles from the International Classification of Diseases, Tenth Revision, chapter E (ICD-E).

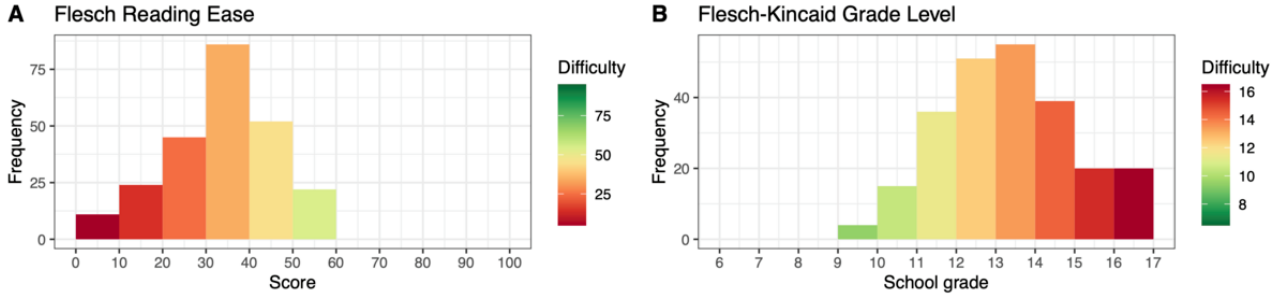
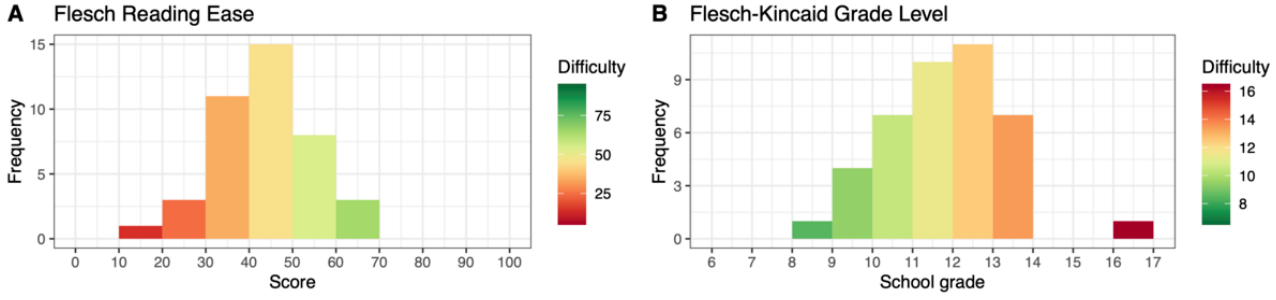


Figure 7. Distribution of Flesch Reading Ease and Flesch-Kincaid Grade Level values for Russian articles from the International Classification of Diseases, Tenth Revision, chapter O (ICD-O).



XSL•FO
RenderX

Table 3. Mean and SD of each readability metric for groups A and B for articles from English Wikipedia (N=6127).

Groups	Values, n (%)	Flesch Reading Ease, mean (SD)	Flesch-Kincaid Grade Level, mean (SD)	Automated Readability Index, mean (SD)	Coleman-Liau Index, mean (SD)	Simple Measure of Gobbledygook, mean (SD)	Gunning Frequency of Gobbledygook, mean (SD)
A[a]	4233 (69.11)	28.69 (11.00)	14.26 (1.80)	13.65 (0.85)	15.16 (1.08)	16.12 (1.53)	16.70 (0.78)
B	1892 (30.89)	29.18 (12.86)	14.47 (2.11)	13.57 (1.10)	14.87 (1.38)	16.33 (1.84)	16.61 (0.99)

[a]In group A, 2 articles were excluded from the analysis because of technical inability to compute the readability metrics.

Table 4. Mean and SD of each readability metric for each International Classification of Diseases, Tenth Revision, chapter individually for articles from English Wikipedia (N=4471).

International Classification of Diseases chapters	Values, n (%)	Flesch Reading Ease, mean (SD)	Flesch-Kincaid Grade Level, mean (SD)	Automated Readability Index, mean (SD)	Coleman-Liau Index, mean (SD)	Simple Measure of Gobbledygook, mean (SD)	Gunning Frequency of Gobbledygook, mean (SD)
A	107 (2.39)	32.63 (9.25)	13.69 (1.59)	13.67 (0.70)	15.04 (0.87)	15.71 (1.40)	16.60 (1.00)
B	95 (2.12)	33.24 (9.07)	13.53 (1.51)	13.55 (0.83)	14.80 (1.00)	15.60 (1.24)	16.64 (0.76)
C	216 (4.83)	29.91 (9.80)	14.03 (1.69)	13.66 (0.86)	15.20 (1.05)	15.80 (1.43)	16.66 (0.80)
D	281 (6.28)	27.67 (11.72)	14.28 (1.86)	13.58 (0.98)	15.17 (1.11)	16.07 (1.61)	16.65 (0.88)
E	414 (9.26)	25.14 (9.88)	14.81 (1.53)	13.83 (0.61)	15.48 (0.78)	16.44 (1.37)	16.87 (0.50)
F	342 (7.65)	23.88 (9.95)	15.33 (1.53)	13.87 (0.51)	15.49 (0.85)	17.22 (1.38)	16.92 (0.42)
G	324 (7.25)	27.90 (10.02)	14.57 (1.64)	13.78 (0.67)	15.21 (0.97)	16.37 (1.38)	16.80 (0.56)
H	239 (5.35)	30.04 (10.57)	13.93 (1.85)	13.48 (1.02)	14.99 (1.14)	15.86 (1.54)	16.64 (0.79)
I	183 (4.09)	30.05 (9.07)	14.29 (1.61)	13.77 (0.68)	15.25 (1.04)	16.20 (1.30)	16.79 (0.62)
J	118 (2.64)	29.79 (11.03)	13.94 (1.93)	13.50 (0.92)	15.22 (1.01)	15.75 (1.52)	16.52 (0.87)
K	200 (4.47)	27.48 (10.08)	14.37 (1.60)	13.76 (0.78)	15.37 (0.96)	16.20 (1.34)	16.79 (0.63)
L	192 (4.29)	29.38 (13.61)	13.79 (2.04)	13.40 (1.03)	15.04 (1.30)	15.85 (1.67)	16.54 (0.99)
M	258 (5.77)	31.17 (11.94)	13.67 (1.91)	13.44 (1.21)	15.02 (1.31)	15.59 (1.51)	16.52 (1.02)
N	158 (3.53)	30.11 (9.36)	13.93 (1.73)	13.61 (0.82)	15.31 (0.94)	15.79 (1.48)	16.61 (0.85)
O	85 (1.9)	30.21 (9.99)	14.24 (1.54)	13.72 (0.61)	14.98 (1.11)	16.36 (1.33)	16.82 (0.51)
P	73 (1.63)	30.22 (10.30)	14.30 (1.62)	13.75 (0.83)	14.95 (1.31)	16.27 (1.26)	16.79 (0.56)
Q	740 (16.55)	28.05 (11.95)	14.24 (1.83)	13.60 (0.86)	15.10 (1.10)	16.10 (1.57)	16.70 (0.76)
R	228 (5.1)	28.73 (10.54)	14.41 (1.74)	13.73 (0.72)	15.18 (1.00)	16.21 (1.45)	16.75 (0.72)
S	79 (1.77)	39.02 (8.22)	12.78 (1.58)	13.22 (1.03)	14.06 (1.15)	14.66 (1.34)	15.86 (1.29)
T	111 (2.48)	32.06 (10.51)	14.06 (1.86)	13.62 (0.85)	14.78 (1.12)	15.98 (1.51)	16.62 (0.94)
U	5 (0.11)	25.66 (5.18)	15.77 (0.77)	14.00 (0.00)	15.57 (0.48)	17.37 (0.66)	17.00 (0.00)
W	4 (0.09)	43.50 (9.89)	12.57 (1.26)	12.87 (0.80)	12.81 (2.62)	14.26 (1.38)	15.58 (1.05)
X	5 (0.11)	29.60 (8.49)	14.50 (1.45)	13.77 (0.51)	15.17 (1.12)	15.87 (1.10)	16.67 (0.75)
Y	5 (0.11)	28.41 (8.92)	15.01 (1.66)	14.00 (0.00)	15.20 (0.89)	16.73 (1.37)	16.89 (0.24)
Z	9 (0.2)	28.90 (7.43)	14.92 (1.14)	14.00 (0.00)	15.39 (0.99)	16.56 (0.87)	17.00 (0.00)

Table 5. Mean and SD of each readability metric for groups A and B for articles from German Wikipedia (N=6024).

Groups	Values, n (%)	Flesch Reading Ease, mean (SD)	Fourth Vienna Formula, mean (SD)
A	4625 (76.8)	20.33 (9.98)	13.43 (1.28)
B[a]	1397 (23.2)	23.91 (11.49)	13.05 (1.52)

[a]In group B, 2 articles were excluded from the analysis because of technical inability to compute the readability metrics.

XSL•FO
RenderX

Table 6. Mean and SD of each readability metric for each International Classification of Diseases, Tenth Revision, chapter individually for articles from German Wikipedia (N=5092).

International Classification of Diseases chapters	Values, n (%)	Flesch Reading Ease, mean (SD)	Fourth Vienna Formula, mean (SD)
A	112 (2.2)	22.91 (8.77)	13.10 (1.22)
B	99 (1.94)	23.71 (8.14)	13.01 (1.24)
C	227 (4.46)	19.83 (9.26)	13.45 (1.30)
D	311 (6.11)	18.54 (10.22)	13.52 (1.28)
E	416 (8.17)	17.43 (9.80)	13.59 (1.17)
F	312 (6.13)	23.10 (9.50)	13.38 (1.34)
G	408 (8.01)	20.07 (10.51)	13.43 (1.29)
H	282 (5.54)	21.91 (10.06)	13.23 (1.33)
I	202 (3.97)	18.60 (8.91)	13.62 (1.13)
J	118 (2.32)	19.90 (9.39)	13.53 (1.32)
K	224 (4.4)	21.08 (9.95)	13.27 (1.31)
L	206 (4.05)	21.25 (10.15)	13.30 (1.31)
M	297 (5.83)	19.80 (9.30)	13.48 (1.16)
N	173 (3.4)	18.70 (9.09)	13.56 (1.23)
O	74 (1.45)	22.61 (9.94)	13.37 (1.23)
P	65 (1.28)	19.79 (8.59)	13.60 (1.16)
Q	1030 (20.23)	19.25 (10.60)	13.54 (1.29)
R	268 (5.26)	20.37 (9.54)	13.61 (1.26)
S	107 (2.1)	24.60 (8.22)	13.06 (1.30)
T	130 (2.55)	23.27 (9.55)	13.31 (1.24)
U	6 (0.12)	20.76 (6.51)	13.93 (1.11)
W	5 (0.1)	24.97 (7.90)	13.34 (1.30)
X	3 (0.06)	26.47 (5.86)	12.83 (0.83)
Y	5 (0.1)	23.01 (14.15)	13.42 (2.16)
Z	12 (0.24)	17.15 (9.97)	13.76 (1.10)

Table 7. Mean and SD of each readability metric for groups A and B for articles from Russian Wikipedia (N=3314).

Groups	Values, n (%)	Flesch Reading Ease, mean (SD)	Flesch-Kincaid Grade Level, mean (SD)
A	2316 (69.91)	38.54 (13.51)	12.64 (1.84)
B[a]	997 (30.09)	38.82 (15.34)	12.58 (2.10)

[a]In group B, 1 article was excluded from the analysis because of technical inability to compute the readability metrics.

XSL•FO
RenderX

Table 8. Mean and SD of each readability metric for each International Classification of Diseases, Tenth Revision, chapter individually for articles from Russian Wikipedia (N=2417).

International Classification of Diseases chapters	Values, n (%)	Flesch Reading Ease, mean (SD)	Flesch-Kincaid Grade Level, mean (SD)
A	83 (3.43)	42.03 (11.38)	12.16 (1.50)
B	78 (3.23)	42.22 (9.22)	12.16 (1.32)
C	105 (4.34)	36.10 (13.01)	12.86 (1.75)
D	102 (4.22)	36.07 (13.58)	13.01 (1.83)
E	240 (9.93)	33.66 (12.73)	13.35 (1.73)
F	275 (11.38)	35.68 (14.78)	13.02 (2.02)
G	146 (6.04)	37.43 (12.76)	12.80 (1.72)
H	186 (7.7)	41.43 (13.59)	12.23 (1.85)
I	117 (4.84)	35.55 (13.69)	13.04 (1.76)
J	75 (3.1)	37.99 (12.26)	12.73 (1.69)
K	115 (4.76)	39.95 (12.03)	12.47 (1.68)
L	88 (3.64)	41.99 (12.36)	12.18 (1.72)
M	119 (4.92)	39.04 (12.72)	12.58 (1.72)
N	100 (4.14)	36.43 (12.43)	12.97 (1.75)
O	41 (1.7)	44.06 (10.73)	11.88 (1.53)
P	34 (1.41)	37.41 (14.67)	12.78 (2.00)
Q	234 (9.68)	41.52 (14.43)	12.21 (1.96)
R	162 (6.7)	39.12 (14.17)	12.55 (1.93)
S	29 (1.1.2)	42.97 (14.90)	11.99 (2.00)
T	76 (3.14)	42.87 (11.96)	12.04 (1.68)
U	5 (0.21)	37.52 (7.76)	12.77 (1.10)
W	1 (0.04)	55.33 (—[a])	10.18 (—)
X	2 (0.08)	46.07 (1.43)	11.58 (0.24)
Y	1 (0.04)	14.07 (—)	16.16 (—)
Z	3 (0.12)	32.72 (8.61)	13.51 (1.14)

[a]No SD could be computed for 1 article.

Further descriptive values, such as minimum and maximum values of each readability metric, as well as mean (SD) and minimum and maximum values of sentence and word and complex word count for each ICD-10 chapter, can be found in Multimedia Appendix 10.

Comparison of Readability

Comparison Among Languages

Interlanguage comparisons were conducted pairwise between each resulting pair of the 3 languages. For this purpose, the FRE values of all articles (ie, groups A and B) were considered. The readability of the Wikipedia texts (English: FRE=28.84; German: FRE=21.16; Russian: FRE=38.62) differed significantly among these values. The results of the corresponding unpaired 1-tailed *t* tests are presented in Table 9.

The distribution of articles in English, German, and Russian Wikipedia based on their difficulty is depicted in Figure 8. The frequency of articles in Figure 8 is lower than that of the other 2 subplots, as the total number of articles from Russian Wikipedia is smaller than that from English and German Wikipedia.

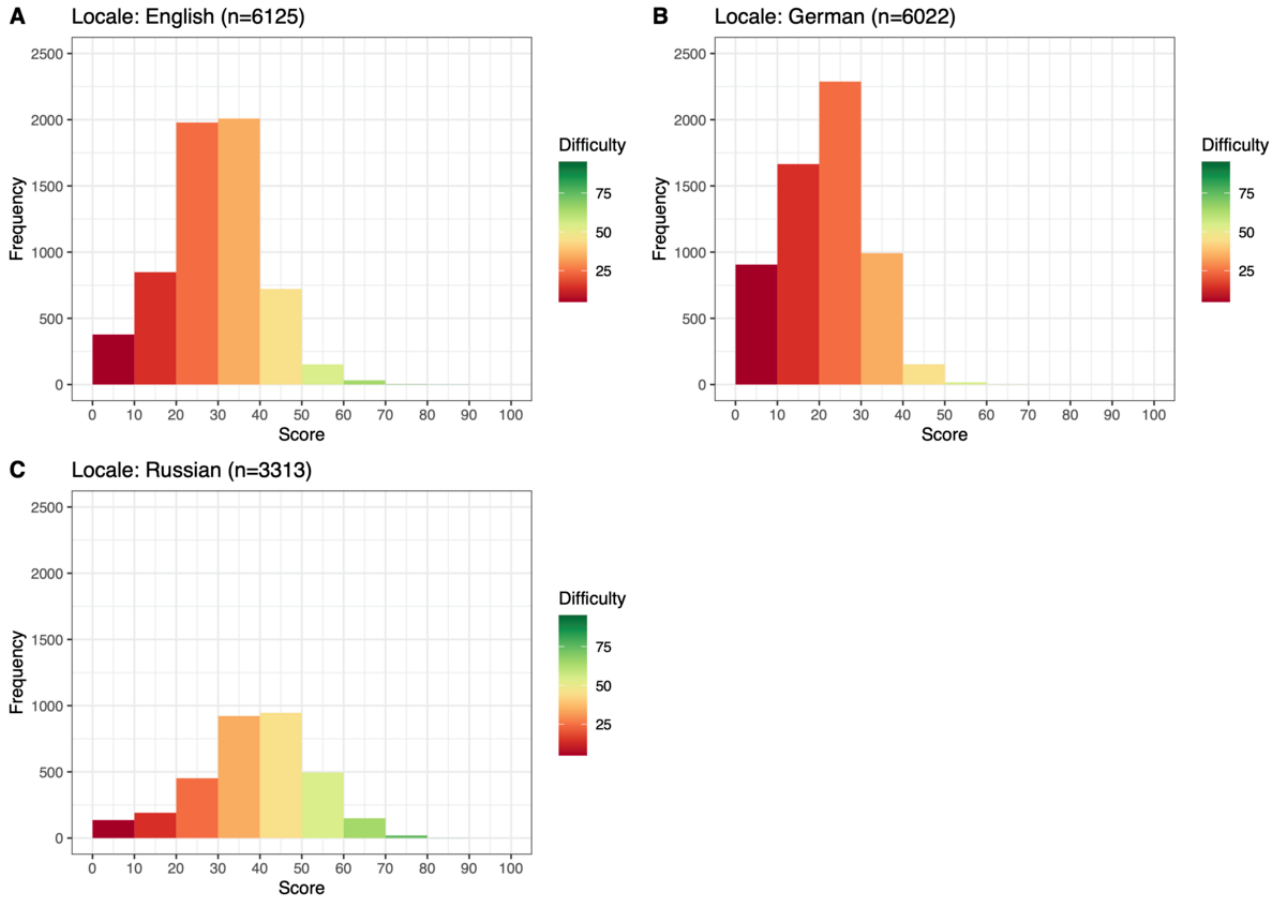
More than 90% (9/10) of English Wikipedia articles (5937/6125, 96.93%) had an FRE value <50.00, which signals that they are difficult to extremely difficult to read. In total, 99.7% (6004/6022) of articles from German Wikipedia have an FRE value <50.00 and fall under the same category. Among the articles included in the analysis from Russian Wikipedia, 79.9% (2647/3313) had an FRE value <50.00.

XSL•FO
RenderX

Table 9. Comparison of Wikipedia articles in different languages for readability difficulty.

Comparison	Difference of means (95% CI)	*P* value
English vs German	7.68 (7.29 to 8.08)	<.001
English vs Russian	−9.78 (−10.34 to −9.22)	<.001
German vs Russian	−17.46 (−18.01 to −16.92)	<.001

Figure 8. Distribution of Flesch Reading Ease (FRE) values of all articles included in the analysis from English (A), German (B), and Russian (C) Wikipedia.



Comparison Among ICD-10 Chapters

Unpaired 2-tailed *t* tests were conducted to investigate differences among ICD-10 chapters compared with the average FRE value of group A articles. This analysis was performed separately for each language. The average FRE values of group A articles were 28.69 (SD 11.00), 20.33 (SD 9.98), and 38.54 (SD 13.51) for English, German, and Russian Wikipedia, respectively (Tables 3, 5, and 7). Figure 9, Figure 10, and Figure 11 depict median and mean FRE values of every ICD-10 chapter for English, German, and Russian, respectively.

All values are below the FRE score of 70.00, which means that no articles in English group A were easy to read. Articles with the highest FRE scores required at least 8 or 9 years of education for adequate text comprehension.

Significant differences were found for 9 ICD-10 chapters in articles collected from English Wikipedia: chapter F (FRE 23.88, SD 9.95; *P*<.001), chapter E (FRE 25.14, SD 9.88; *P*<.001), chapter H (FRE 30.04, SD 10.57; *P*=.049), chapter I (FRE 30.05, SD 9.07; *P*=.04), chapter M (FRE 31.17, SD 11.94; *P*<.001), chapter T (FRE 32.06, SD 10.51; *P*=.001), chapter A (FRE 32.63, SD 9.25; *P*<.001), chapter B (FRE 33.24, SD 9.07; *P*<.001), and chapter S (FRE 39.02, SD 8.22; *P*<.001).

For the German group A, all articles scored below the FRE of 55.00 and, thus, were at least *fairly difficult* to *extremely difficult* to read. To adequately understand these articles, the reader needs at least 10 years of prior education.

In articles collected from German Wikipedia, the average readability values of 11 ICD-10 chapters differed significantly from the average value of group A articles: chapter E (FRE 17.43, SD 9.80; *P*<.001), chapter D (FRE 18.54, SD 10.22; *P*=.002), chapter I (FRE 18.60, SD 8.91; *P*=.006), chapter N (FRE 18.70, SD 9.09; *P*=.02), chapter Q (FRE 19.25, SD 10.60; *P*=.001), chapter H (FRE 21.91, SD 10.06; *P*=.009), chapter A (FRE 22.91, SD 8.77; *P*=.002), chapter F (FRE 23.10, SD 9.50; *P*<.001), chapter T (FRE 23.27, SD 9.55; *P*<.001), chapter B (FRE 23.71, SD 8.14; *P*<.001), and chapter S (FRE 24.60, SD 8.22; *P*<.001).

XSL•FO
RenderX

In total, 0.26% (6/2316) of group A articles from Russian Wikipedia had an FRE score between 70.00 and 80.00; that is, they were fairly easy to read and required 7 years of education for text comprehension.

For Russian Wikipedia articles, 10 ICD-10 chapters showed a significantly different difficulty in comparison with the group A average value: chapter E (FRE 33.66, SD 12.73; $P<.001$), chapter I (FRE 35.55, SD 13.69; $P=.02$), chapter F (FRE 35.68, SD 14.78; $P=.001$), chapter H (FRE 41.43, SD 13.59; $P=.004$), chapter Q (FRE 41.52, SD 14.43; $P=.002$), chapter L (FRE 41.99, SD 12.36; $P=.01$), chapter A (FRE 42.03, SD 11.38; $P=.006$), chapter B (FRE 42.22, SD 9.22; $P<.001$), chapter T (FRE 42.87, SD 11.96; $P=.002$), and chapter O (FRE 44.06, SD 10.73; $P=.002$).

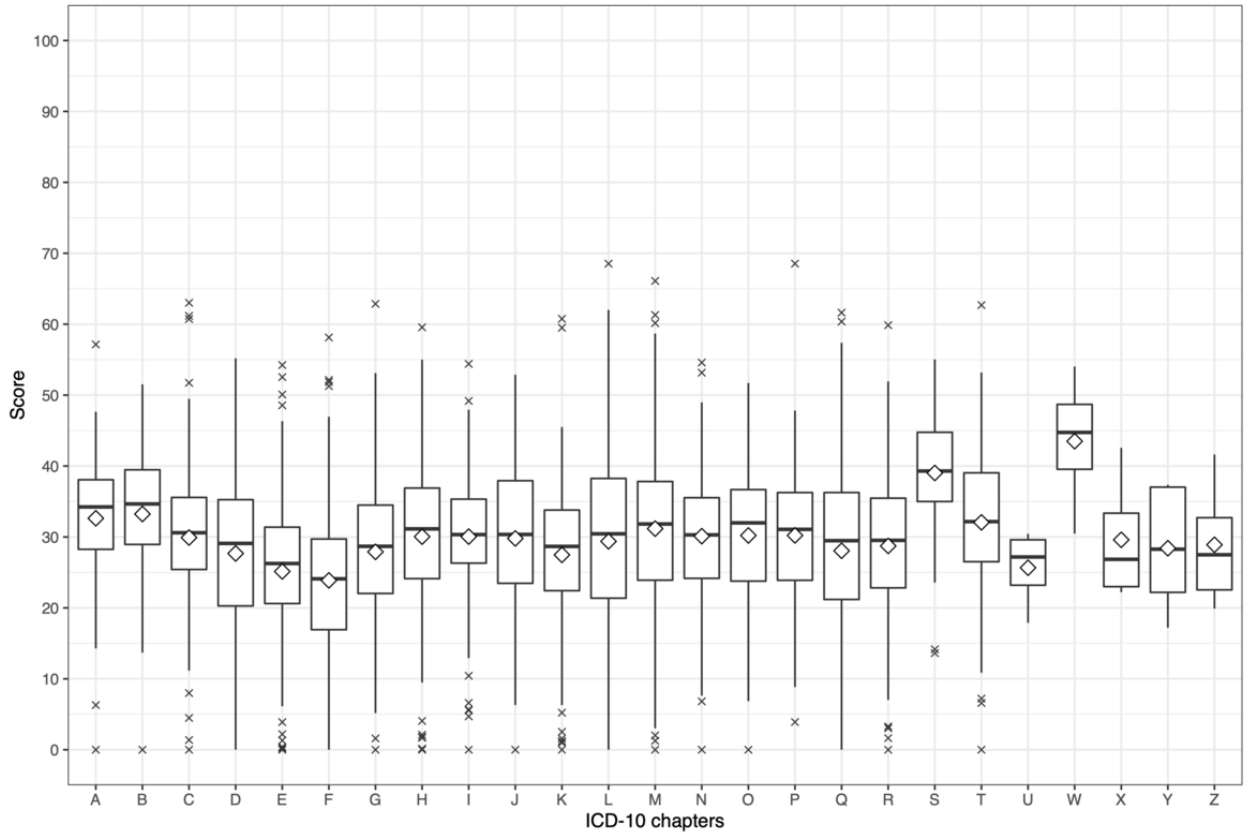
For English and German, articles associated with the ICD-10 chapter S were, on average, significantly easier to read than all articles from group A. Articles from Russian Wikipedia were, on average, easier to read but showed no significant difference from the mean value. Chapter S codes represent *injury, poisoning, and certain other consequences of external causes.*

In contrast, articles associated with ICD-10 chapter F, which codes *mental, behavioral, and neurodevelopmental disorders*, were significantly harder to understand than an average group A article from English and Russian Wikipedia. The respective average value of this chapter in German Wikipedia was even lower than that in English Wikipedia but significantly higher than the average value of group A articles for German locale. Thus, in all 3 Wikipedia domains, articles related to mental, behavioral, and neurodevelopmental disorders were, on average, very difficult to read. For adequate comprehension, readers would need at least a college degree.

Chapter B (*certain infectious and parasitic diseases*), chapter T (*injury, poisoning, and certain other consequences of external causes*), chapter I (*diseases of the circulatory system*), and chapter H (*diseases of eye and adnexa*) had significant differences with the average readability of group A articles in all 3 languages. In Russian and German languages, chapters B, T, and H had significantly higher FRE values than the average, and chapter I had lower FRE values. For the data collection of English articles, all 4 chapters had significantly better readability than average.

Details for every ICD-10 chapter and each language can be found in Multimedia Appendix 10.

Figure 9. Box plot depicting Flesch Reading Ease values in each International Classification of Diseases, Tenth Revision (ICD-10), chapter for articles from English Wikipedia. The rhombus represents the mean Flesch Reading Ease value.



XSL•FO
RenderX

Figure 10. Box plot depicting Flesch Reading Ease values in each International Classification of Diseases, Tenth Revision (ICD-10) chapter for articles from German Wikipedia. The rhombus represents the mean Flesch Reading Ease value.

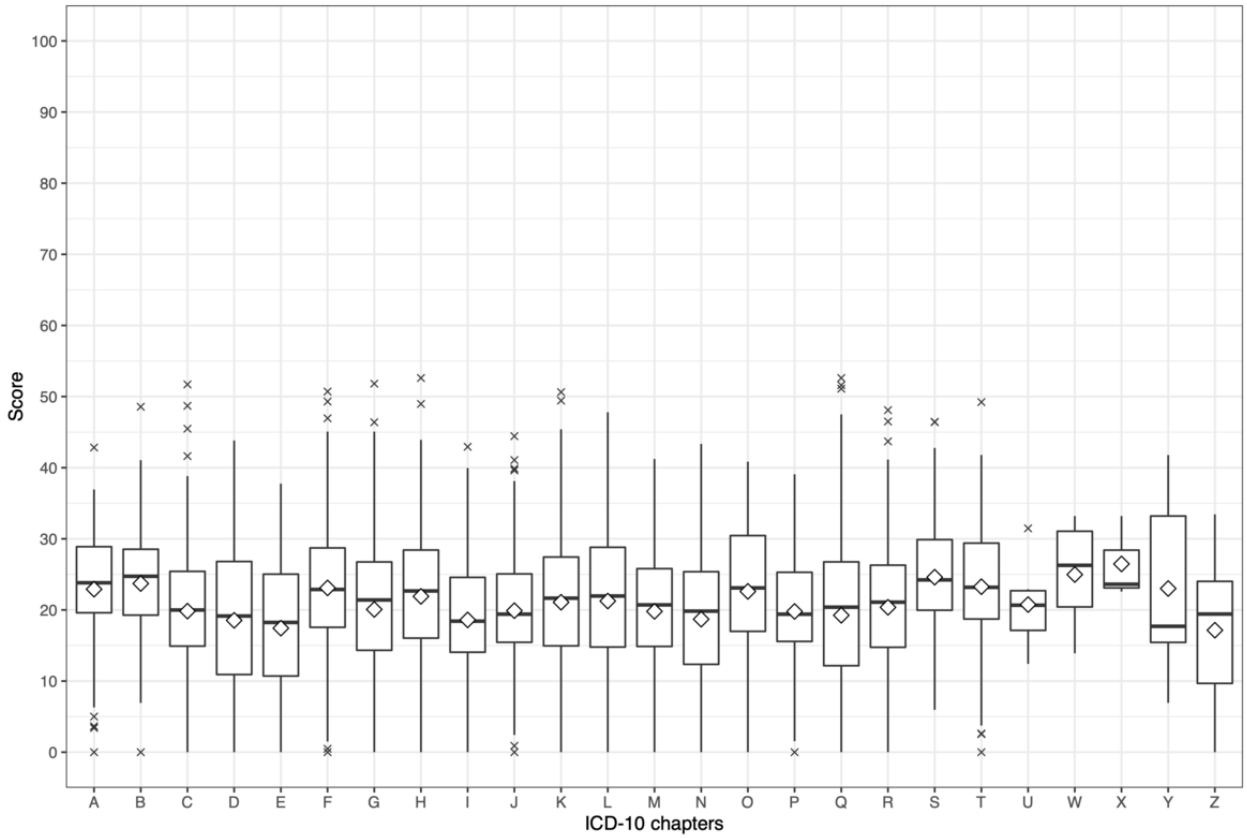
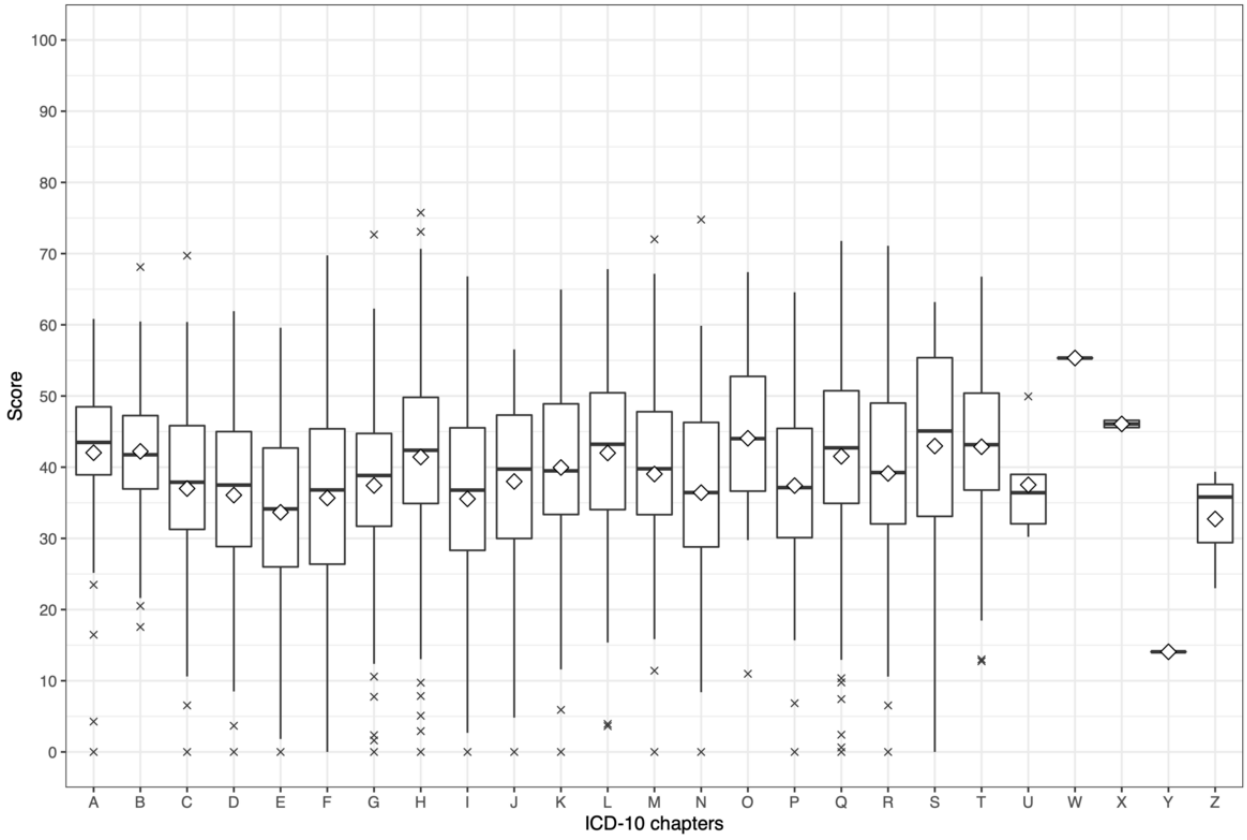


Figure 11. Box plot depicting Flesch Reading Ease values in each International Classification of Diseases, Tenth Revision (ICD-10), chapter for articles from Russian Wikipedia. The rhombus represents the mean Flesch Reading Ease value.



XSL•FO
RenderX

Comparison With Recommended Grade Level

For the English language, there is a recommended readability level of roughly *7-8* [26]. An unpaired 1-tailed *t* test was performed for articles from English Wikipedia. On average, every ICD-10 chapter had a significantly higher FKGL than the recommended grade level of 7 ($P<.001$).

Further details on FKGL results can be found in Multimedia Appendix 11.

Discussion

Principal Findings

Most articles were collected from English Wikipedia. Moreover, the original graph from English Wikipedia had the highest number of nodes in comparison with graphs generated from Russian and German Wikipedia. However, only 31.78% (1947/6127) of the articles included for the English locale were acquired in graph-based data collection (processing step 1). The remaining articles were collected in processing step 2 based on cross-linking from German and Russian Wikipedia. A possible reason for this may be the unsuitable category structure of English Wikipedia, which prevents the construction of an optimal graph for further processing. In contrast, German Wikipedia and Russian Wikipedia were found to have a better categorical structure, which led to most articles from these localizations being collected in the first graph-based data collection step.

Furthermore, Wikipedia articles provided good ICD-10 code coverage for disease-related articles. Most articles (English: 4235/6127, 69.12%; German: 4625/6024, 76.78%; Russian: 2316/3314, 69.89%) were assigned to group A because an ICD-10 code was provided.

The presence of an ICD-10 code associated with an article in Wikipedia may encourage readers to gather further information about the disease they are interested in. However, Wikipedia is not a medical textbook, and annotating ICD codes to an article is not strictly required. Moreover, it is questionable whether layperson readers or writers would be familiar with the concept of ICD codes. Although readers with a medical professional background could make use of it, the question remains whether those people are the target audience of Wikipedia.

On average, English articles in every single ICD-10 chapter failed to meet the readability of recommendation of the computed grade level of 7 [26].

In general, disease-related articles from all 3 Wikipedia domains were difficult and very difficult to read: pages from Russian Wikipedia were significantly easier to read with an overall FRE score of 38.62 (difficult to read), followed by pages from English (FRE 28.84) and German (FRE 21.16) Wikipedia with very difficult-to-understand texts. The relatively easier-to-understand ICD-10 chapter from German Wikipedia was still very difficult to read (FRE 24.60). On average, the least difficult chapters in English (FRE 39.02) and Russian (FRE 44.06) were identified as difficult to read. Most of the analyzed articles required a college to a professional level of education for adequate comprehension. Therefore, disease-related articles found on Wikipedia cannot be recommended as stand-alone educational materials for patients seeking information on the web. Nevertheless, patients will be confronted with this, as they use search engines that present Wikipedia pages in the top ranks [8-11].

Limitations

One of the general limitations is the choice of the readability metric. In this study, classical readability metrics based on sentence and word structures were used for the analysis. However, such readability metrics provide insight into only one aspect of the understandability of a given text. Readability formulas ignore factors that can contribute to ease of reading but are not based on sentence structure and word length (eg, illustrations, sentence connection, and syntax). The role of the reader was also not taken into consideration [68]. Several studies have shown that other approaches to the analysis of readability have higher concordance with human assessment than the established readability formulas [45,46].

Moreover, the chosen readability metrics were mainly based on the ASL and number of syllables in each word, as well as language-specific weighting factors. However, estimating the number of syllables in a word is not a trivial task for German and Russian languages and does not always work reliably [69]. For this reason, the computed values could be affected by inaccuracies in the syllable count. In this context, it should be stressed that this affects all natural language processing analysis tools for German and Russian text materials.

There are also several technical limitations to be mentioned in the context of the study. First, Wikipedia graphs were generated based on data exported from June 2021. Therefore, our findings are valid only for that moment in time, as Wikipedia is updated daily and exports from June 2021 may not reflect the current situation.

Second, some relevant articles could have been missed because of the choice of the main concept and category filtering for each language. Similarly, some irrelevant articles could have been included in the analysis if they were incorrectly categorized by Wikipedia authors. However, irrelevant articles were mostly found in group B because their ICD-10 codes were not specified.

Furthermore, Wikipedia pages can be created by anyone and might not be strictly reviewed before publishing or making changes. Thus, inaccuracies and inconsistencies might be the outcomes for the readers of a page. In this context, the ICD-10 code could have been specified incorrectly. Some of the retrieved ICD-10 codes were manually excluded or adjusted by the authors. For instance, an article from German Wikipedia—*Computerspielabhängigkeit* (English: video game addiction; Russian: Зависимость от компьютерных игор)—was retrieved with an ICD-10 code *C51*, which stands for malignant neoplasms. In this case, the ICD-10 code template was used incorrectly by the authors of the article, who instead assigned an ICD-11 code (6C51 gaming disorder) to this article. There were also articles with falsely retrieved ICD-10 codes; for example, *0.00* was retrieved instead of *Q0.00* for the English article *Anencephaly* (German: *Anenzephalie*; Russian: Анэнцефалия).

XSL•FO
RenderX

Comparison With Prior Work

Previous studies have investigated readability for a relatively small sample of disease- or health-related English Wikipedia pages [10,18,19,34-36]. All studies consistently found that these are difficult to read and require at least a college degree or higher level of education. Our study confirms that English Wikipedia pages (still) require a college graduate degree on average.

Kräenbring et al [44] investigated the readability of German Wikipedia pages related to pharmacology. They found that, on average, related articles had a WSTF value of 15.04. In our study, disease-related pages in German showed an average WSTF value of 13.43 (group A, SD 1.28) and 13.05 (group B, SD 1.52). Our findings confirm that higher education is necessary to understand German Wikipedia articles.

To the best of the authors' knowledge, this study is the first to analyze the disease-related content of Russian Wikipedia. No further studies have investigated the level of readability in comparison with other languages. As reported for English and German localizations, our findings indicate that readers of Russian articles require a comparatively high level of education.

In contrast to previous studies, this study analyzed the readability of 3 different Wikipedia languages, with each sample containing thousands of articles. In addition, this study presents a detailed comparison of all medical subfields based on the internationally adopted ICD-10 classification. All details of the analyses are available in Multimedia Appendices 3-11.

Future Directions

In future work, the readability of other popular languages such as Spanish, French, and Chinese could be investigated to check whether differences exist. Furthermore, vocabulary analysis of Wikipedia pages can be conducted to add another dimension of understandability to established readability metrics, as demonstrated in the study by Zowalla et al [69].

Moreover, a consecutive study could draw a direct comparison of simplified versions of Wikipedia, such as *Simple English* (available on the web [70]). In this context, readability levels are expected to be significantly easier than those of regular versions.

Although readability analysis provides valuable insights into the understandability of texts, an investigation of content quality (eg, using DISCERN) could be beneficial for assessing the suitability of Wikipedia articles as educational material for patients. Similarly, the analysis of visual elements and information depictions could allow a more thorough investigation in the context of this topic. However, visual interpretation is subject to personal preferences, varies substantially, and requires tremendous manual effort. From a technical perspective, this is a challenge for even modern image-processing libraries.

To increase the readability of articles, Wikipedia can provide a built-in readability check that can serve two purposes: (1) informing authors about the readability values of the text and, thus, encouraging them to provide easier formulations and descriptions and (2) informing the reader about the difficulty of the currently displayed article.

Conclusions

For the English, German, and Russian editions, disease-related Wikipedia pages were difficult to read and understand. For adequate comprehension, a college degree is required to understand articles from Russian Wikipedia, and a graduate degree is required for readers of the English and German Wikipedia editions.

Therefore, Wikipedia in all 3 languages cannot be recommended as a stand-alone patient education material. It does not meet the recommended readability for such materials written in English. Although no such recommendations are available for German and Russian, our findings confirm the low readability of pages for all 3 Wikipedia localizations.

The authors of Wikipedia pages should carefully revise existing text materials for readers with a specific interest in a disease or its associated symptoms. Special attention should be given to articles on *mental, behavioral, and neurodevelopmental disorders* (ICD-10 chapter F) as these articles were most difficult to read (FKGL 15.33, ARI 13.87, CLI 15.49, SMOG 17.22, and Gunning FOG 16.92) in comparison with other ICD-10 chapters.

A built-in readability indicator could be useful for authors contributing to Wikipedia. This would increase readability at the text production stage and, thus, allow more people to comprehend medical knowledge through encyclopedias, which are freely available on the internet.

Authors' Contributions

MW and MP conceptualized the study. MW and RZ implemented the technical software components. JG performed the literature review and readability analyses and drafted the first version of the manuscript. All the authors revised and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematic literature review.
[PDF File (Adobe PDF File), 345 KB - jmir_v24i5e36835_app1.pdf]

XSL•FO
RenderX

Multimedia Appendix 2
General statistics for disease-related Wikipedia graphs.
[PDF File (Adobe PDF File), 53 KB - jmir_v24i5e36835_app2.pdf]

Multimedia Appendix 3
Terms and wildcards applied for category filtering.
[PDF File (Adobe PDF File), 57 KB - jmir_v24i5e36835_app3.pdf]

Multimedia Appendix 4
Distributions of all computed readability metrics in English.
[PDF File (Adobe PDF File), 2227 KB - jmir_v24i5e36835_app4.pdf]

Multimedia Appendix 5
Distributions of all computed readability metrics in German.
[PDF File (Adobe PDF File), 1263 KB - jmir_v24i5e36835_app5.pdf]

Multimedia Appendix 6
Distributions of all computed readability metrics in Russian.
[PDF File (Adobe PDF File), 2025 KB - jmir_v24i5e36835_app6.pdf]

Multimedia Appendix 7
Box plots with readability values for the English sample.
[PDF File (Adobe PDF File), 522 KB - jmir_v24i5e36835_app7.pdf]

Multimedia Appendix 8
Box plots with readability values for the German sample.
[PDF File (Adobe PDF File), 143 KB - jmir_v24i5e36835_app8.pdf]

Multimedia Appendix 9
Box plots with readability values for the Russian sample.
[PDF File (Adobe PDF File), 174 KB - jmir_v24i5e36835_app9.pdf]

Multimedia Appendix 10
Tables containing mean (SD) and minimum and maximum values for all computed readability metrics for articles in groups A and B, as well as for every International Classification of Diseases, Tenth Revision, chapter separately.
[XLSX File (Microsoft Excel File), 39 KB - jmir_v24i5e36835_app10.xlsx]

Multimedia Appendix 11
P and respective mean values of all conducted t tests for the second and third test scenarios.
[XLSX File (Microsoft Excel File), 17 KB - jmir_v24i5e36835_app11.xlsx]

References

1. Marton C, Wei Choo C. A review of theoretical models of health information seeking on the web. J Doc 2012 Apr 20;68(3):330-352. [doi: 10.1108/00220411211225575]
2. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
3. Keinki C, Zowalla R, Wiesner M, Koester MJ, Huebner J. Understandability of patient information booklets for patients with cancer. J Cancer Educ 2018 Jun;33(3):517-527. [doi: 10.1007/s13187-016-1121-3] [Medline: 27726109]
4. Sillence E, Briggs P, Harris PR, Fishwick L. How do patients evaluate and make use of online health information? Soc Sci Med 2007 May;64(9):1853-1862. [doi: 10.1016/j.socscimed.2007.01.012] [Medline: 17328998]
5. Moreland J, French TL, Cumming GP. The prevalence of online health information seeking among patients in Scotland: a cross-sectional exploratory study. JMIR Res Protoc 2015 Jul 15;4(3):e85 [FREE Full text] [doi: 10.2196/resprot.4010] [Medline: 26177562]
6. Bjelke M, Martinsson AK, Lendahls L, Oscarsson M. Using the Internet as a source of information during pregnancy - a descriptive cross-sectional study in Sweden. Midwifery 2016 Sep;40:187-191. [doi: 10.1016/j.midw.2016.06.020] [Medline: 27450590]

XSL•FO
RenderX

7. Shea-Budgell MA, Kostaras X, Myhill KP, Hagen NA. Information needs and sources of information for patients during cancer follow-up. Curr Oncol 2014 Aug;21(4):165-173 [FREE Full text] [doi: 10.3747/co.21.1932] [Medline: 25089098]
8. The top 500 sites on the web. Alexa. 2021. URL: https://www.alexa.com/topsites [accessed 2021-12-16]
9. Top-500 Registered Domains of the Latest Main Crawl. Common Crawl. URL: https://commoncrawl.github.io/cc-crawl-statistics/plots/domains [accessed 2021-12-16]
10. Hutchinson N, Baird GL, Garg M. Examining the reading level of Internet medical information for common internal medicine diagnoses. Am J Med 2016 Jun;129(6):637-639. [doi: 10.1016/j.amjmed.2016.01.008] [Medline: 26829438]
11. Seth AK, Vargas CR, Chuang DJ, Lee BT. Readability assessment of patient information about lymphedema and its treatment. Plast Reconstr Surg 2016 Feb;137(2):287e-295e. [doi: 10.1097/01.prs.0000475747.95096.ab] [Medline: 26818318]
12. Wikipedia: size of Wikipedia. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia [accessed 2021-12-16]
13. Chrzanowski J, Sołek J, Fendler W, Jemielniak D. Assessing public interest based on Wikipedia's most visited medical articles during the SARS-CoV-2 outbreak: search trends analysis. J Med Internet Res 2021 Apr 12;23(4):e26331 [FREE Full text] [doi: 10.2196/26331] [Medline: 33667176]
14. Matheson D, Matheson-Monnet C. Wikipedia as informal self-education for clinical decision-making in medical practice. Open Med J 2017 Sep 30;4(Supplement 1):15-25. [doi: 10.2174/1874220301704010015]
15. Smith DA. Situating Wikipedia as a health information resource in various contexts: a scoping review. PLoS One 2020 Feb 18;15(2):e0228786 [FREE Full text] [doi: 10.1371/journal.pone.0228786] [Medline: 32069322]
16. Farič N, Potts HW. Motivations for contributing to health-related articles on Wikipedia: an interview study. J Med Internet Res 2014 Dec 03;16(12):e260 [FREE Full text] [doi: 10.2196/jmir.3569] [Medline: 25498308]
17. Volsky PG, Baldassari CM, Mushti S, Derkay CS. Quality of Internet information in pediatric otolaryngology: a comparison of three most referenced websites. Int J Pediatr Otorhinolaryngol 2012 Sep;76(9):1312-1316. [doi: 10.1016/j.ijporl.2012.05.026] [Medline: 22770592]
18. John AM, John ES, Hansberry DR, Thomas PJ, Guo S. Analysis of online patient education materials in pediatric ophthalmology. J AAPOS 2015 Oct;19(5):430-434. [doi: 10.1016/j.jaapos.2015.07.286] [Medline: 26486024]
19. Shetty KR, Wang RY, Shetty A, Levi J, Aaronson NL. Quality of patient education sections on Otitis media across different website platforms. Ann Otol Rhinol Laryngol 2020 Jun;129(6):591-598. [doi: 10.1177/0003489420902183] [Medline: 31975608]
20. Garcia SF, Hahn EA, Jacobs EA. Addressing low literacy and health literacy in clinical oncology practice. J Support Oncol 2010;8(2):64-69 [FREE Full text] [Medline: 20464884]
21. Alpay L, Verhoef J, Xie B, Te'eni D, Zwetsloot-Schonk JH. Current challenge in consumer health informatics: bridging the gap between access to information and information understanding. Biomed Inform Insights 2009 Jan 01;2(1):1-10 [FREE Full text] [doi: 10.4137/bii.s2223] [Medline: 20419038]
22. 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] [Medline: 15626525]
23. Broom A. Virtually he@lthy: the impact of Internet use on disease experience and the doctor-patient relationship. Qual Health Res 2005 Mar;15(3):325-345. [doi: 10.1177/1049732304272916] [Medline: 15761103]
24. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
25. Oh HJ, Lee B. The effect of computer-mediated social support in online communities on patient empowerment and doctor-patient communication. Health Commun 2012;27(1):30-41. [doi: 10.1080/10410236.2011.567449] [Medline: 21797714]
26. How to Write Easy-to-Read Health Materials. US National Library of Medicine. URL: https://web.archive.org/web/20180519094225/https://medlineplus.gov/etr.html [accessed 2019-08-28]
27. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
28. Gunning R. The Technique of Clear Writing. New York, NY, USA: McGraw Hill; 1968.
29. McLaughlin GH. SMOG grading: a new readability formula. J Read 1969;12(8):639-646.
30. Coleman M, Liau TL. A computer readability formula designed for machine scoring. J Appl Psychol 1975;60(2):283-284. [doi: 10.1037/h0076540]
31. Fry E. Fry's readability graph: clarifications, validity, and extension to level 17. J Read 1977 Dec;21(3):242-252.
32. Amstad T. Wie verständlich sind unsere Zeitungen?. Zürich, Switzerland: Studenten-Schreib-Service; 1978.
33. Solovyev V, Ivanov V, Solnyshkina M. Assessment of reading difficulty levels in Russian academic texts: approaches and metrics. J Intell Fuzzy Syst 2018 May 24;34(5):3049-3058. [doi: 10.3233/jifs-169489]
34. Suwannakhan A, Casanova-Martínez D, Yurasakpong L, Montriwat P, Meemon K, Limpanuparb T. The quality and readability of English Wikipedia anatomy articles. Anat Sci Educ 2020 Jul;13(4):475-487. [doi: 10.1002/ase.1910] [Medline: 31233658]

XSL•FO
RenderX

35. Brigo F, Otte WM, Igwe SC, Tezzon F, Nardone R. Clearly written, easily comprehended? The readability of websites providing information on epilepsy. Epilepsy Behav 2015 Mar;44:35-39. [doi: 10.1016/j.yebeh.2014.12.029] [Medline: 25601720]
36. Handler SJ, Eckhardt SE, Takashima Y, Jackson AM, Truong C, Yazdany T. Readability and quality of Wikipedia articles on pelvic floor disorders. Int Urogynecol J 2021 Dec;32(12):3249-3258. [doi: 10.1007/s00192-021-04776-0] [Medline: 33797592]
37. McEnteggart GE, Naeem M, Skierkowski D, Baird GL, Ahn SH, Soares G. Readability of online patient education materials related to IR. J Vasc Interv Radiol 2015 Aug;26(8):1164-1168. [doi: 10.1016/j.jvir.2015.03.019] [Medline: 25935147]
38. McInnes N, Haglund BJ. Readability of online health information: implications for health literacy. Inform Health Soc Care 2011 Dec;36(4):173-189. [doi: 10.3109/17538157.2010.542529] [Medline: 21332302]
39. Miles RC, Baird GL, Choi P, Falomo E, Dibble EH, Garg M. Readability of online patient educational materials related to breast lesions requiring surgery. Radiology 2019 Apr;291(1):112-118. [doi: 10.1148/radiol.2019182082] [Medline: 30694156]
40. Modiri O, Guha D, Alotaibi NM, Ibrahim GM, Lipsman N, Fallah A. Readability and quality of Wikipedia pages on neurosurgical topics. Clin Neurol Neurosurg 2018 Mar;166:66-70. [doi: 10.1016/j.clineuro.2018.01.021] [Medline: 29408776]
41. Punia V, Dagar A, Agarwal N, He W, Hillen M. Comparison of neurological healthcare oriented educational resources for patients on the Internet. J Clin Neurosci 2014 Dec;21(12):2179-2183. [doi: 10.1016/j.jocn.2014.05.043] [Medline: 25194822]
42. Rajagopalan MS, Khanna VK, Leiter Y, Stott M, Showalter TN, Dicker AP, et al. Patient-oriented cancer information on the Internet: a comparison of Wikipedia and a professionally maintained database. J Oncol Pract 2011 Sep;7(5):319-323 [FREE Full text] [doi: 10.1200/JOP.2010.000209] [Medline: 22211130]
43. Tulbert BH, Snyder CW, Brodell RT. Readability of patient-oriented online dermatology resources. J Clin Aesthet Dermatol 2011 Mar;4(3):27-33 [FREE Full text] [Medline: 21464884]
44. Kräenbring J, Monzon Penza T, Gutmann J, Muehlich S, Zolk O, Wojnowski L, et al. Accuracy and completeness of drug information in Wikipedia: a comparison with standard textbooks of pharmacology. PLoS One 2014 Sep 24;9(9):e106930 [FREE Full text] [doi: 10.1371/journal.pone.0106930] [Medline: 25250889]
45. Zheng J, Yu H. Assessing the readability of medical documents: a ranking approach. JMIR Med Inform 2018 Mar 23;6(1):e17 [FREE Full text] [doi: 10.2196/medinform.8611] [Medline: 29572199]
46. Kauchak D, Leroy G, Hogue A. Measuring text difficulty using parse-tree frequency. J Assoc Inf Sci Technol 2017 Sep;68(9):2088-2100 [FREE Full text] [doi: 10.1002/asi.23855] [Medline: 29057293]
47. Readability Formulas. URL: https://readabilityformulas.com/ [accessed 2022-05-04]
48. Wikipedia - The Free Encyclopedia. URL: https://en.wikipedia.org/wiki/Main_Page [accessed 2022-05-04]
49. Willkommen bei Wikipedia. URL: https://de.wikipedia.org/wiki/Wikipedia:Hauptseite [accessed 2022-05-04]
50. Добро пожаловать в Википедию. URL: https://ru.wikipedia.org/wiki/%D0%97%D0%B0%D0%B3%D0%BB%D0%B0%D0%B2%D0%BD%D0%B0%D1%8F%D1%81%D1%82%D1%80%D0%B0%D0%BD%D0%B8%D1%86%D0%B0 [accessed 2022-05-04]
51. ICD-10: International statistical classification of diseases and related health problems: 10th revision. World Health Organization. 2016. URL: https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf [accessed 2022-03-22]
52. Frequently asked questions. World Health Organization. URL: https://www.who.int/standards/classifications/frequently-asked-questions [accessed 2022-03-22]
53. Template:ICD-10. Wikipedia, the free encyclopedia. 2022. URL: https://en.wikipedia.org/wiki/Template:ICD10 [accessed 2022-03-22]
54. Vorlage:Infobox ICD. Wikipedia, the free encyclopedia. 2022. URL: https://de.wikipedia.org/wiki/Vorlage:Infobox_ICD [accessed 2022-03-22]
55. Шаблон:ICD10. Wikipedia, the free encyclopedia. 2022. URL: https://ru.wikipedia.org/wiki/%D0%A8%D0%B0%D0%B1%D0%BB%D0%BE%D0%BD:ICD10 [accessed 2022-03-22]
56. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inform Decis Mak 2021 Nov 09;21(Suppl 6):206 [FREE Full text] [doi: 10.1186/s12911-021-01534-6] [Medline: 34753471]
57. International Statistical Classification of Diseases and Related Health Problems (ICD). World Health Organization. 2021. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2021-12-20]
58. Klare GR. Assessing readability. Read Res Q 1974;10(1):62-102. [doi: 10.2307/747086]
59. Klare GR. The formative years. In: Zakaluk BL, Samuels SJ, editors. Readability: its past, present, and future. Newark, DE, USA: International Reading Association; 1988:14-34.
60. Kincaid JP, Fishburne Jr RP, Rogers RL, Chissom BS. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Institute for Simulation and Training. 1975. URL: https://stars.library.ucf.edu/istlibrary/56/ [accessed 2022-05-03]
61. Smith EA, Senter RJ. Automated readability index. AMRL TR 1967 May:1-14. [Medline: 5302480]
62. Bamberger R, Vanecek E. Lesen - Verstehen - Lernen - Schreiben: die Schwierigkeitsstufen von Texten in deutscher Sprache. Wien, Austria: Jugend und Volk; 1984.

XSL•FO
RenderX

63. Wikimedia Downloads. URL: https://dumps.wikimedia.org/ [accessed 2022-05-04]
64. Wikidata Toolkit. MediaWiki. URL: https://www.mediawiki.org/wiki/Wikidata_Toolkit [accessed 2022-05-03]
65. Wiesner M, Zowalla R, Pobiruchin M. The difficulty of German information booklets on psoriasis and psoriatic arthritis: automated readability and vocabulary analysis. JMIR Dermatol 2020 Feb 28;3(1):e16095. [doi: 10.2196/16095]
66. Wickham H. ggplot2: Elegant Graphics For Data Analysis. 2nd edition. New York, NY, USA: Springer; 2016.
67. Holm S. A simple sequentially rejective multiple test procedure. Scand J Stat 1979;6(2):65-70. [doi: 10.2307/4615733]
68. Using readability formulas: a cautionary note. Centers for Medicare & Medicaid Services, U.S. Department of Health and Human Services. 2010. URL: https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/Downloads/ToolkitPart07.pdf [accessed 2021-12-16]
69. Zowalla R, Wiesner M, Pfeifer D. Automatically assessing the expert degree of online health content using SVMs. Stud Health Technol Inform 2014;202:48-51. [Medline: 25000012]
70. Simple English Wikipedia. URL: https://simple.wikipedia.org/wiki/Main_Page [accessed 2022-05-04]

Abbreviations

ARI: Automated Readability Index
ASL: average sentence length
ASW: average syllables per word
CLI: Coleman-Liau Index
FKGL: Flesch-Kincaid Grade Level
FOG: Frequency of Gobbledygook
FRE: Flesch Reading Ease
ICD-10: International Classification of Diseases, Tenth Revision
ICD-11: International Classification of Diseases, Eleventh Revision
MS: proportion of (complex) words with ≥3 syllables
SMOG: Simple Measure of Gobbledygook
WSTF: Wiener SachTextFormel

Edited by A Mavragani; submitted 27.01.22; peer-reviewed by A Chang; comments to author 26.02.22; revised version received 12.04.22; accepted 14.04.22; published 16.05.22.

Please cite as:
Gordejeva J, Zowalla R, Pobiruchin M, Wiesner M
Readability of English, German, and Russian Disease-Related Wikipedia Pages: Automated Computational Analysis
J Med Internet Res 2022;24(5):e36835
URL: https://www.jmir.org/2022/5/e36835
doi:10.2196/36835
PMID:35576562

©Jelizaveta Gordejeva, Richard Zowalla, Monika Pobiruchin, Martin Wiesner. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Effectiveness of a Computer-Tailored Web-Based Physical Activity Intervention Using Fitbit Activity Trackers in Older Adults (Active for Life): Randomized Controlled Trial

Stephanie J Alley[1], PhD; Jannique van Uffelen[2], PhD; Stephanie Schoeppe[1], PhD; Lynne Parkinson[3], PhD; Susan Hunt[4], PhD; Deborah Power[1], BBus; Natasha Waterman[1]; Courtney Waterman[1]; Quyen G To[1], PhD; Mitch J Duncan[3], PhD; Anthony Schneiders[5], PhD; Corneel Vandelanotte[1], PhD

[1]Physical Activity Research Group, Appleton Institute, Central Queensland University, Rockhampton, Australia

[2]Department of Movement Sciences, KU Leuven, Leuven, Belgium

[3]School of Medicine & Public Health, College of Health, Medicine, and Wellbeing, The University of Newcastle, Callaghan, Australia

[4]School of Nursing, Midwifery and Social Sciences, Central Queensland University, Melbourne, Australia

[5]School of Health, Medical and Applied Sciences, Central Queensland University, Bundaberg, Australia

Corresponding Author:
Stephanie J Alley, PhD
Physical Activity Research Group
Appleton Institute
Central Queensland University
Building 7, Bruce Hwy
Rockhampton, 4701
Australia
Phone: 61 749232263
Email: s.alley@cqu.edu.au

Abstract

Background: Physical activity is an integral part of healthy aging; yet, most adults aged ≥65 years are not sufficiently active. Preliminary evidence suggests that web-based interventions with computer-tailored advice and Fitbit activity trackers may be well suited for older adults.

Objective: The aim of this study was to examine the effectiveness of *Active for Life*, a 12-week web-based physical activity intervention with 6 web-based modules of computer-tailored advice to increase physical activity in older Australians.

Methods: Participants were recruited both through the web and offline and were randomly assigned to 1 of 3 trial arms: tailoring+Fitbit, tailoring only, or a wait-list control. The computer-tailored advice was based on either participants' Fitbit data (tailoring+Fitbit participants) or self-reported physical activity (tailoring-only participants). The main outcome was change in wrist-worn accelerometer (ActiGraph GT9X)–measured moderate to vigorous physical activity (MVPA) from baseline to after the intervention (week 12). The secondary outcomes were change in self-reported physical activity measured by means of the Active Australia Survey at the midintervention point (6 weeks), after the intervention (week 12), and at follow-up (week 24). Participants had a face-to-face meeting at baseline for a demonstration of the intervention and at baseline and week 12 to return the accelerometers. Generalized linear mixed model analyses were conducted with a γ distribution and log link to compare MVPA and self-reported physical activity changes over time within each trial arm and between each of the trial arms.

Results: A total of 243 participants were randomly assigned to tailoring+Fitbit (n=78, 32.1%), tailoring only (n=96, 39.5%), and wait-list control (n=69, 28.4%). Attrition was 28.8% (70/243) at 6 weeks, 31.7% (77/243) at 12 weeks, and 35.4% (86/243) at 24 weeks. No significant overall time by group interaction was observed for MVPA (*P*=.05). There were no significant within-group changes for MVPA over time in the tailoring+Fitbit group (+3%, 95% CI –24% to 40%) or the tailoring-only group (–4%, 95% CI –24% to 30%); however, a significant decline was seen in the control group (–35%, 95% CI –52% to –11%). The tailoring+Fitbit group participants increased their MVPA 59% (95% CI 6%-138%) more than those in the control group. A significant time by group interaction was observed for self-reported physical activity (*P*=.02). All groups increased their self-reported physical activity from baseline to week 6, week 12, and week 24, and this increase was greater in the tailoring+Fitbit group than in the control group at 6 weeks (+61%, 95% CI 11%-133%).

XSL•FO
RenderX

Conclusions: A computer-tailored physical activity intervention with Fitbit integration resulted in improved MVPA outcomes in comparison with a control group in older adults.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618000646246; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12618000646246

(J Med Internet Res 2022;24(5):e31352) doi:10.2196/31352

KEYWORDS

internet; online; activity trackers; activity monitors; wearables; physical activity; mobile phone

Introduction

Background

Physical activity is important for healthy aging. Physical activity improves health and well-being and reduces the risk of chronic disease [1,2]. Older adults who are physically active have improved mobility, a reduced risk of falls, and a reduced risk of cognitive decline [1]. However, <30% of older adults are meeting the physical activity recommendations of 30 minutes of at least moderate-intensity physical activity on most days [3]. These low levels of physical activity are contributing to Australia's rising health costs from the aging population [4-6]. Therefore, population-based physical activity interventions with a wide reach are required.

Web-based physical activity interventions are effective in young and middle-aged adults [7], and they may be well suited to older adults (aged ≥65 years). The percentage of older adults using the internet is steadily growing. In 2016, 79% of older Australians were already connected, of whom 85% used the internet daily [8]. Reviews have found web-based physical activity interventions to be effective in older adults [9-11]. However, many of the included studies used existing interventions created for middle-aged adults rather than new interventions specifically developed for older adults [12,13] and included participants as young as 50 years of age [14,15].

Tailored web-based interventions that provide automated personalized physical activity advice based on participants' characteristics, physical activity, motivation, and environment are effective and may be particularly suited to older adults [16]. This is because older adults have greater diversity of health-related characteristics [3] and because they have expressed the need for physical activity advice to be tailored specifically to them [17]. This expressed need is in line with findings from Ammann et al [18], who found that a tailored web-based physical activity intervention was more effective in older participants than in younger ones. However, very few studies have tested the effectiveness of web-based computer-tailored physical activity interventions created for adults aged ≥65 years and those that have done so demonstrated mixed results [9,11,19,20].

Tailored web-based interventions typically provide participants with personalized advice based on self-reported physical activity data; as such, it is possible that inaccurate advice is delivered because of overreporting of physical activity and social desirability bias [21]. However, commercial activity trackers (eg, Fitbit) allow tailored advice to be based on objectively measured physical activity [22]. A study conducted by Vandelanotte et al [22] found physical activity advice based on Fitbit data to be more credible and lead to greater physical activity changes than advice based on self-reported physical activity data in middle-aged adults. Although older adults do not use activity trackers as frequently as younger adults [23], use is growing steadily in this population [24]. Moreover, multiple studies have found face-to-face, telephone, SMS text messaging, and email advice based on activity tracker data to be effective in older adults [9,25]. Most of these interventions were conducted in older adults with a specific chronic illness (patients with cardiac conditions, chronic obstructive pulmonary disease, or osteoarthritis) and gave feedback on walking and step counts [25]. A recent trial in older adults with no chronic illnesses found a combined web and face-to-face intervention based on Fitbit data to be effective [26]. However, no studies in older adults have investigated the effectiveness of fully automated computer-tailored physical activity advice based on physical activity behavior recorded through a Fitbit activity tracker [27].

Objectives

The primary aim of the study was to test the effectiveness of a web-based computer-tailored physical activity intervention with Fitbit (Google LLC) integration at increasing objectively measured moderate to vigorous physical activity (MVPA) from before to after the intervention compared with a web-based computer-tailored physical activity intervention without Fitbit integration and a control group in adults aged ≥65 years. The secondary aims were to compare the web-based computer-tailored intervention with and without Fitbit integration and the control group on objectively measured sedentary behavior from before to after the intervention and to compare subjectively measured physical activity and sitting time changes at the midintervention point, after the intervention, and at follow-up.

We hypothesized that the web-based computer-tailored physical activity intervention with Fitbit integration would lead to increased objectively measured MVPA and self-reported physical activity over time and decreased objectively measured sedentary behavior and self-reported sitting time over time compared with the web-based computer-tailored physical activity intervention without Fitbit integration and a control group.

XSL•FO
RenderX

Methods

Study Design

A 3-arm randomized controlled trial was conducted where participants were randomized into one of three groups: (1) tailoring+Fitbit, (2) tailored advice only, and (3) wait-list control. Participants completed web-based surveys at baseline (week 0), at the midintervention point (week 6), after the intervention (week 12), and at follow-up (week 24). Objective physical activity and sedentary behavior were collected by means of wrist-worn accelerometry at baseline and week 12. More detail of the methods can be found in a protocol paper of the trial [28].

Participants

Participants were recruited in Rockhampton (regional Queensland), Bundaberg (regional Queensland), and Adelaide (metropolitan South Australia), Australia, through paid Facebook advertising, email lists, flyers, and local newsletters. Recruitment was carried out between April 2018 and March 2019, and data collection was completed in November 2019. Eligible participants were English-speaking adults aged ≥65 years who had internet access and basic internet confidence, could attend 2 face-to-face appointments at one of the project locations, and could safely increase their physical activity as determined by the Physical Activity Readiness Questionnaire [29] or general physician approval. Eligible participants were those not meeting the physical activity guidelines [30], as assessed by asking participants the following question: "Are you currently participating in less than 30 minutes of physical activity on 5 days a week?" Participants were ineligible if they were already participating in another physical activity program or had used a Fitbit activity tracker in the previous 6 months.

Sample Size Analysis

To detect differences between the 2 intervention groups and the control group for accelerometer-measured MVPA from baseline to after the intervention, 100 participants per group were required. This was to detect an effect size of 0.37 based on the average effect size of web-based physical activity interventions for inactive adults [31]. This accounted for a dropout rate of up to 30%. Power was set at 0.80 and the α at .05. The decision to end participant recruitment was made by the lead (SJA) and senior (CV) investigators because the trial was close to the sample size goal (n=243), the remaining funds were limited, and interest in the trial had slowed.

Procedures

Advertising materials directed prospective participants to the landing page of the intervention website, which had more details about study participation and access to the participant information sheet and eligibility survey. Prospective participants were automatically notified of their eligibility upon completion of the survey, and eligible participants received a welcome email. Participants were asked to complete web-based research surveys at baseline, week 6, week 12, and week 24 through the intervention website. Participants indicated their informed consent through a check box at the beginning of the baseline survey. If participants missed a survey, they were still asked to complete later surveys. Participants were posted an accelerometer to wear on their wrist for 7 consecutive 24-hour days, including when sleeping and showering, at baseline and week 12. The blinded accelerometers were only used for research evaluation and were not part of the intervention. Participants attended a baseline appointment to return the baseline accelerometer and were randomly allocated to one of the three trial arms (tailoring+Fitbit, tailored advice only, and wait-list control). Randomization lists were created by the lead investigator (SJA) using computer-automated block randomization with block sizes of 15 and a 1:1:1 ratio. Randomization was stratified by sex (male and female) and age (<75 years and >75 years) to ensure an equal distribution of men and women in different age groups over the intervention arms. Using the randomization lists, the research manager (DP) and research assistants (CW and NW) assigned the participants by date of baseline appointment. Because of the nature of the intervention, neither the researchers nor the participants were blinded to group allocation. During the appointment, after randomization, intervention group participants were shown through the *Active for Life* intervention website and Fitbit participants were provided with a Fitbit activity tracker and shown how to sync it to the intervention website. After the 12-week intervention, participants attended another face-to-face follow-up appointment to return the week 12 accelerometer. Participants received up to 3 email reminders for each research survey. If the surveys were still incomplete after the reminders, participants were offered a voucher worth Aus $20 (US $15) to complete them within the next few days. Participants received a voucher worth Aus $50 (US $37) after completing all research surveys. Wait-list participants were given access to the intervention after completing the week 24 research survey.

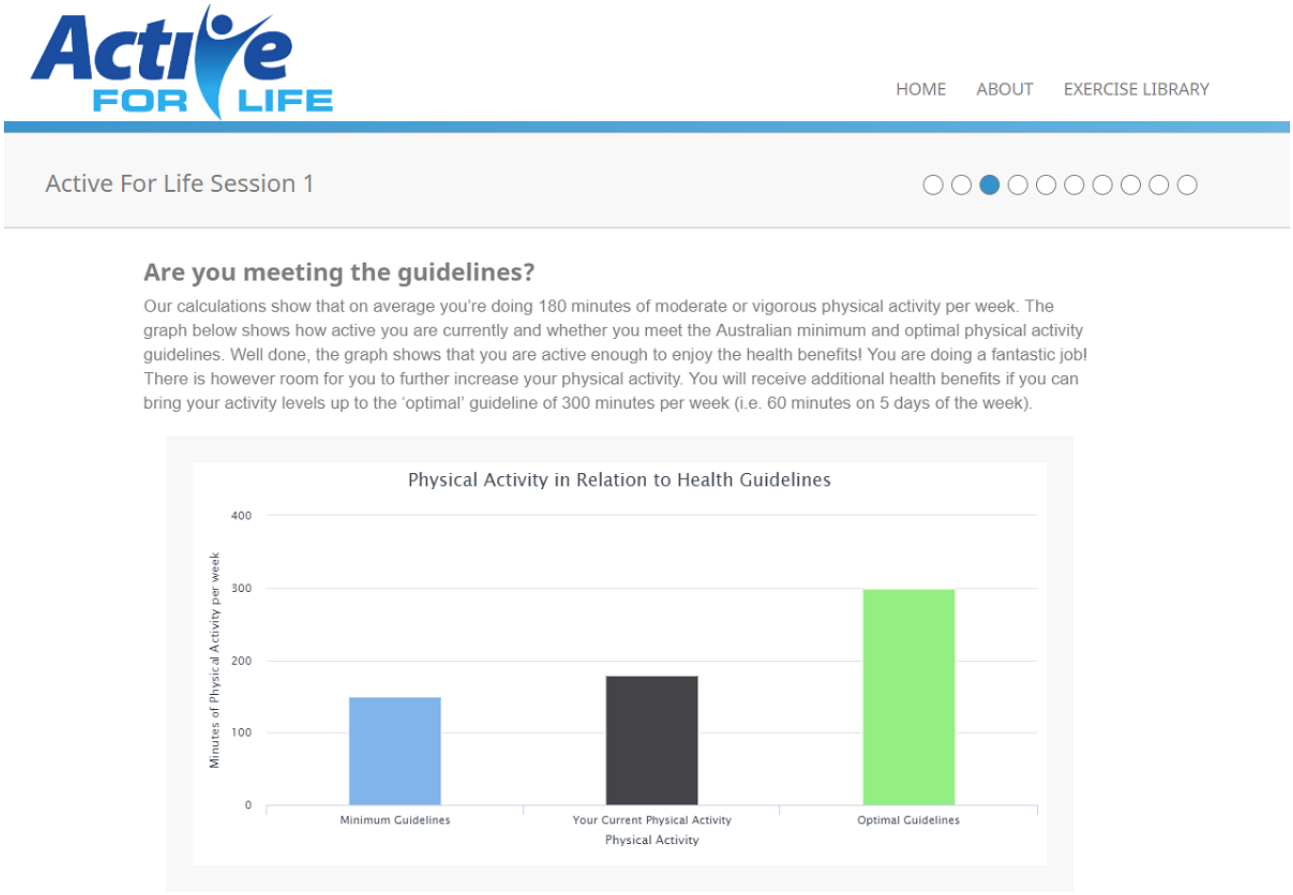
Intervention

An in-depth description of the intervention can be found elsewhere [28]. The *Active for Life* intervention is a 12-week web-based program with 6 modules of tailored advice delivered biweekly. The modules of tailored advice are computer automated and use participant data to select appropriate messages from a database of messages using if-then algorithms (eg, if *low self-efficacy and inactive* then *message on improving self-efficacy by starting small*). The advice is based on the theory of planned behavior [32] and the social cognitive theory [33] and includes evidence-based behavior change techniques [34,35]. The advice encourages participants to work toward meeting the physical recommendations of 30 minutes of moderate-intensity physical activity on at least 5 days each week, including 2 to 3 sessions of strength and flexibility activity. Participants were also encouraged to limit their sitting time to <8 hours per day and to take regular breaks from sitting. Specifically, the advice covers the physical activity recommendations, physical activity benefits, safety when exercising, exercising with a chronic disease, sedentary behavior, goal setting, action plans, self-efficacy, physical activity barriers, social and physical environments, rewards, habit formation, and relapse prevention. The physical activity advice is tailored to participants' characteristics and environment, physical activity behavior, and psychosocial correlates of physical activity (eg, self-efficacy and social

XSL•FO
RenderX

support). Each module of advice included approximately 10 sections that participants scroll through, each with a paragraph on a new subtopic (eg, *Are you meeting the guidelines?*, *Losing weight*, and *Exercise with arthritis*). Some sections include a graph or a picture (Figure 1).

Figure 1. Tailored advice sample.



The intervention website includes an action-planning tool that participants are encouraged to complete at the end of modules 2 and 4. The tool guides participants in setting an action plan (what, where, when, and with whom) for being active in the following fortnight. The intervention website also includes an exercise library where participants can access strength and flexibility exercise plans, written by a physiotherapist, suitable for the beginner and intermediate levels. The plans are 4 weeks in duration, and participants can view videos of the exercises through a link to an external website (Physitrack PLC).

Both the tailoring+Fitbit and the tailoring-only participants were given access to the same intervention, including the 6 modules of computer-tailored advice, action-planning tool, and exercise library. Both groups completed a brief questionnaire at the start of each module to inform the computer-tailored advice. The only difference was that the tailoring+Fitbit participants were required to sync their Fitbit activity tracker with the website to measure their physical activity over the past 2 weeks, whereas the tailoring-only group answered a few additional questions that asked them to recall how many minutes of physical activity they had completed in the past 2 weeks. All other questions (eg, sitting time, self-efficacy, and social support) were identical. Participants in the control group initially only received access to the research surveys, but after completing the week 24 survey, they also received access to the tailored advice modules, action-planning tool, and exercise library.

Measures

Objective measurement of physical activity and sedentary behavior was carried out by means of wrist-worn ActiGraph GT9X accelerometers at baseline and week 12. Accelerations were recorded at 30 Hz. Accelerometer data were processed through ActiLife and valid wear time was set at 16 hours each day on a minimum of 5 days. Nonwear time was assessed based on vector magnitude using the Choi et al [36] algorithm. This algorithm defines nonwear time as 90 consecutive minutes of 0 counts per minute, with 2-minute interruptions allowed. Physical activity behavior was defined using vector-magnitude cut points for older adults as determined by Kamada et al [37]. Sedentary behavior was defined as <2000 counts per minute, light physical activity as 2000 to 8249 counts per minute, and MVPA as ≥8250 counts per minute. Periods of sleep were determined by the Tudor-Locke algorithm [38].

Physical activity was also assessed in all groups at baseline, week 6, week 12, and week 24 by means of the Active Australia Survey [39], which measures time spent walking and in MVPA

XSL•FO
RenderX

over the previous week. Total weekly physical activity is calculated by adding time spent walking, time spent in moderate physical activity, and time spent in vigorous activity doubled to account for the extra energy expenditure. The Active Australia Survey is reliable (intraclass correlation coefficient=0.64) [40] and validated compared with accelerometer-derived MVPA (r=0.35) in older adults [41].

Sitting time was assessed at baseline, week 6, week 12, and week 24 by means of the Workforce Sitting Questionnaire [42]. The questionnaire measures minutes of sitting time per week during work, television viewing, computer use outside work, transport, and other leisure-time activities on work and nonwork days. Daily sitting time on work and nonwork days is calculated by adding sitting time during all activities on work and nonwork days, respectively. Weekly sitting time is calculated by multiplying sitting time on work and nonwork days by the number of days worked and not worked, respectively, and then adding both outcomes. Average daily sitting time is calculated by dividing weekly sitting time by 7. The Workforce Sitting Questionnaire has adequate test-retest reliability (intraclass correlation coefficient=0.46-0.90) and validity compared with accelerometery (women: r=0.22-0.46, men: r=0.18-0.29) for both work and nonwork days [42].

Participant demographics, including sex, age, marital status (single or married or de facto relationship), height and weight (to calculate BMI), English as main language (yes or no), education level (primary, secondary, technical college, or university), employment (full time, part time, or not working), pretax household income (<Aus $41,599 [US $30,658], Aus $41,600 [US $30,659] to Aus $64,999 [US $47,904], Aus $65,000 [US $47,905] to Aus $103,999 [US $76,647], or ≥Aus $104,000 [US $76,648]), and current health diagnosis (yes or no) were measured at baseline. Internet self-efficacy as assessed by means of the valid and reliable Internet Self-efficacy Scale was also assessed at baseline [43]. The Internet Self-efficacy Scale includes 8 items of internet skills on a 7-point Likert scale ranging from 7 (*strongly agree*) to 1 (*strongly disagree*). Items are added together to produce a summary score; higher scores indicate higher internet self-efficacy.

Data Analysis

Analyses of primary and secondary outcomes followed the intention-to-treat principle. Separate generalized linear mixed model analyses were conducted to test the primary outcome of changes in accelerometer-measured MVPA by group and to test the secondary outcomes of changes in self-reported physical activity, accelerometer-measured sedentary behavior, and self-reported sitting time by group. A γ distribution with log link was used for the analyses on accelerometer-measured MVPA and self-reported physical activity because of positively skewed distributions. A normal distribution with identity link was used for the analyses on accelerometer-measured sedentary behavior and self-reported sitting time. Group (tailoring+Fitbit, tailoring only, and wait-list control) by time (baseline and week 12) interactions on accelerometer-measured MVPA and sedentary behavior were analyzed. These analyses controlled for accelerometer wear time. Group (tailoring+Fitbit, tailoring only, and wait-list control) by time (baseline, week 6, week 12, and week 24) interactions on self-reported physical activity and sitting time were analyzed. A sensitivity analysis was conducted to determine the effect of missing data on analysis outcomes. Under the assumption of missing at random, missing values were imputed through chained equations. The fully conditional specification was used to create 20 imputed data sets that were used to conduct the sensitivity analysis. Analyses were conducted using SAS (version 9.4; SAS Institute Inc) with an α of .05.

Ethics Approval

Ethics approval was received from the Central Queensland University Human Ethics Committee before data collection commenced (H16/12-321).

Results

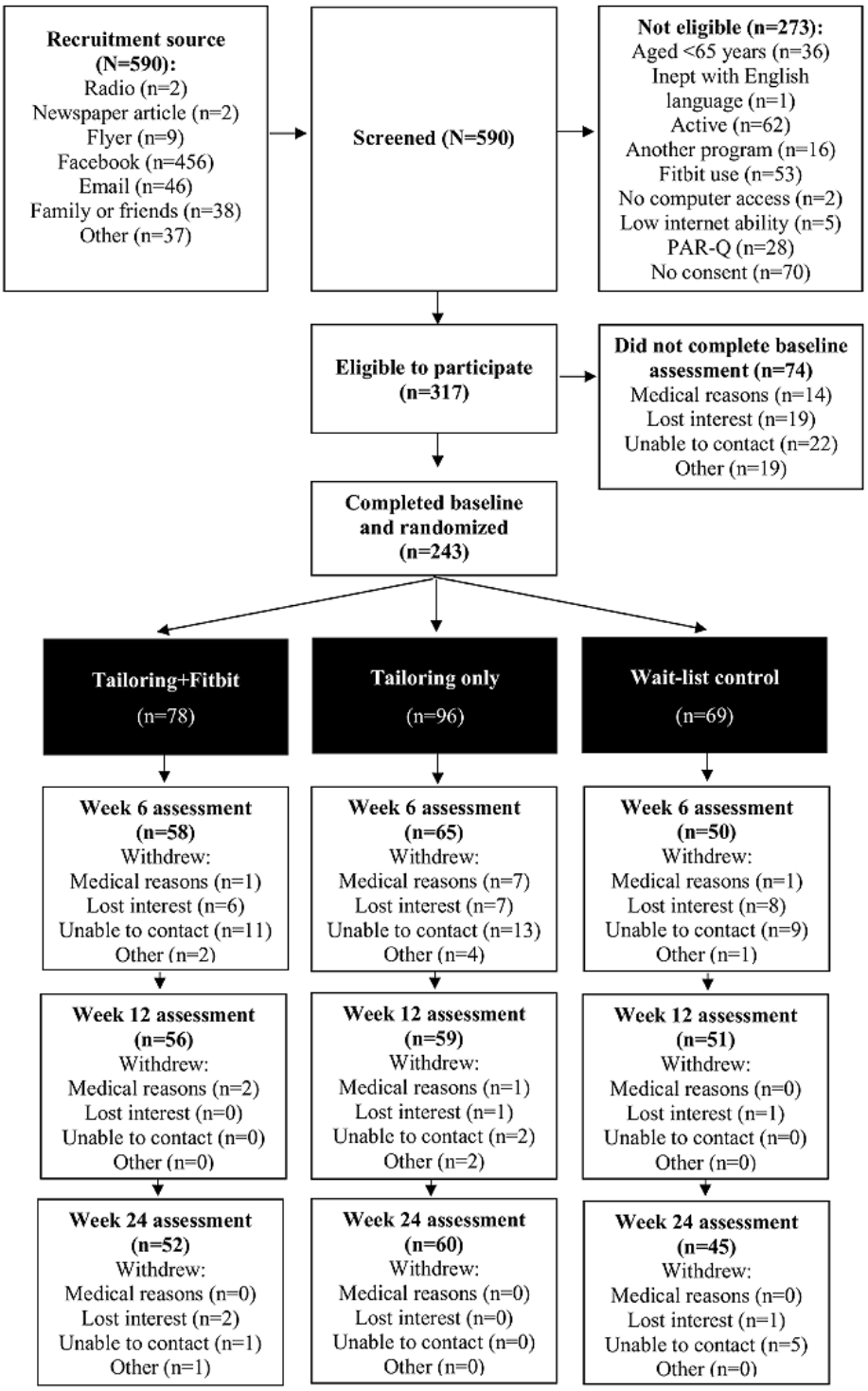
Overview

Figure 2 shows the flow of participants through the trial. Of the 590 participants screened, 317 (53.7%) met the eligibility criteria and 243 (41.2%) completed their baseline assessment and were randomized. Attrition was 28.8% (70/243) at 6 weeks, 31.7% (77/243) at 12 weeks, and 35.4% (86/243) at 24 weeks.

Table 1 shows baseline participant characteristics. Among the 243 participants, 191 (78.6%) were women; 151 (62.1%) were from Adelaide, South Australia; 231 (95.1%) spoke English as their primary language; 172 (70.8%) were married; 126 (51.9%) had a university education; 178 (73.3%) were not working; 89 (36.6%) had a chronic disease; 184 (75.7%) used the internet several times a day; and 84 (44.4%) had a household income <Aus $40,000 (US $29,595). The average age was 69 (SD 4.32; range 65-98) years. The average BMI was 30 (SD 28.84) kg/m^2 (overweight), and the average internet self-efficacy was good at 44 (SD 47.00) out of 56 [43]. Participants who completed the week 12 outcomes had a lower baseline BMI (mean 28.95, SD 5.91, kg/m^2) than participants who did not (mean 31.41, SD 6.37, kg/m^2). No other differences were observed for demographics, health, or internet use between participants who completed the week 12 outcomes and those who did not.

XSL•FO
RenderX

Figure 2. Flowchart of study participants.



XSL•FO
RenderX

Table 1. Baseline participant characteristics (N=243).

Baseline characteristics	All participants	Tailoring+Fitbit (n=78)	Tailoring only (n=96)	Wait-list control (n=76)	Dropout (n=77)	Completer (n=166)	Chi-square (*df*)	*t* test (*df*)	*P* value
Sex, n (%)							0.3 (241)	N/A[a]	.62
Male	52 (21.4)	18 (23.1)	19 (19.8)	15 (21.7)	24 (46.2)	28 (53.8)			
Female	191 (78.6)	60 (76.9)	77 (80.2)	54 (78.3)	78 (40.8)	113 (59.2)			
Location, n (%)							0.6 (241)	N/A	.76
Rockhampton, Queensland	74 (30.5)	24 (30.8)	36 (37.5)	14 (20.3)	24 (32.4)	50 (67.6)			
Bundaberg, Queensland	18 (7.4)	8 (10.3)	5 (5.2)	5 (7.2)	7 (38.9)	11 (61.1)			
Adelaide, South Australia	151 (62.1)	46 (59)	55 (57.3)	50 (72.5)	46 (30.5)	105 (69.5)			
Primary language, n (%)							0.6 (241)	N/A	.53
English	231 (95.1)	72 (92.3)	94 (97.9)	65 (94.2)	72 (31.2)	159 (68.8)			
Other	12 (4.9)	6 (7.7)	2 (2.1)	4 (5.8)	5 (41.7)	7 (58.3)			
Marital status, n (%)							0.02 (241)	N/A	.99
Single	71 (29.2)	22 (28.2)	32 (33.3)	17 (24.6)	22 (31)	49 (69)			
Married or de facto relationship	172 (70.8)	56 (71.8)	64 (66.7)	52 (75.4)	55 (32)	117 (68)			
Education, n (%)							1.8 (241)	N/A	.41
Secondary school	61 (25.1)	25 (32.1)	21 (21.8)	15 (21.7)	19 (31.1)	42 (68.9)			
Technical college	56 (23)	11 (14.1)	29 (29.2)	17 (24.6)	14 (25)	42 (75)			
University	126 (51.9)	42 (53.8)	47 (49)	37 (53.6)	44 (34.9)	82 (65.1)			
Employment, n (%)							2.6 (241)	N/A	.28
Full time	22 (9.1)	8 (10.3)	7 (7.3)	7 (10.1)	7 (31.8)	15 (68.2)			
Part time or casual	43 (17.7)	16 (20.5)	15 (15.7)	12 (17.3)	18 (41.9)	25 (58.1)			
Not working	178 (73.3)	54 (69.2)	74 (77.1)	50 (72.5)	52 (29.2)	126 (70.8)			
Chronic disease status, n (%)							1.2 (241)	N/A	.32
Yes	89 (36.6)	26 (33.3)	35 (36.5)	28 (40.6)	32 (36)	57 (64)			
No	154 (63.4)	52 (66.7)	61 (63.5)	41 (59.4)	45 (29.2)	109 (70.8)			
Internet use, n (%)							1.9 (241)	N/A	.39
Once to several times a week	25 (10.3)	8 (10.3)	13 (13.5)	4 (5.8)	10 (40)	15 (60)			
Once a day	34 (14)	6 (7.7)	18 (18.8)	10 (14.5)	8 (23.5)	26 (76.5)			
Several times a day	184 (75.7)	64 (82.1)	65 (67.7)	55 (79.7)	59 (32.1)	125 (67.9)			
Income,[b] Aus $ (US $), n (%)							2.0 (187)	N/A	.58
>104,000 (>76,957)	24 (12.7)	9 (11.5)	9 (9.4)	6 (8.7)	6 (25)	18 (75)			
65,000 to 103,999 (48,098 to 76,957)	29 (15.3)	10 (12.8)	10 (10.4)	9 (13)	8 (27.6)	21 (72.4)			
41,000 to 64,999 (30,339 to 48,098)	52 (27.5)	11 (14.1)	26 (27.1)	15 (21.7)	19 (36.5)	33 (63.5)			
<40,000 (<29,599)	84 (44.4)	30 (38.5)	33 (34.4)	21 (30.4)	22 (26.2)	62 (73.8)			

XSL•FO
RenderX

Baseline characteristics	All participants	Tailoring+Fitbit (n=78)	Tailoring only (n=96)	Wait-list control (n=76)	Dropout (n=77)	Completer (n=166)	Chi-square (*df*)	*t* test (*df*)	*P* value
Age (years), mean (SD)	69.34 (4.32)	69.88 (4.10)	69.12 (4.93)	68.84 (3.85)	69.08 (3.77)	69.46 (4.56)	N/A	0.64 (241)	.52
BMI,[c] mean (SD)	29.73 (28.84)	29.34 (28.40)	29.46 (28.23)	30.52 (29.74)	31.41 (6.37)	28.95 (5.91)	N/A	2.91 (237)	.004
Internet self-efficacy, mean (SD)	43.90 (47.00)	43.74 (44.50)	44.92 (52.00)	42.65 (45.00)	43.31 (12.43)	44.17 (12.43)	N/A	0.50 (241)	.62

[a]N/A: not applicable.

[b]Income missing (did not wish to disclose): n=54.

[c]BMI missing: n=4.

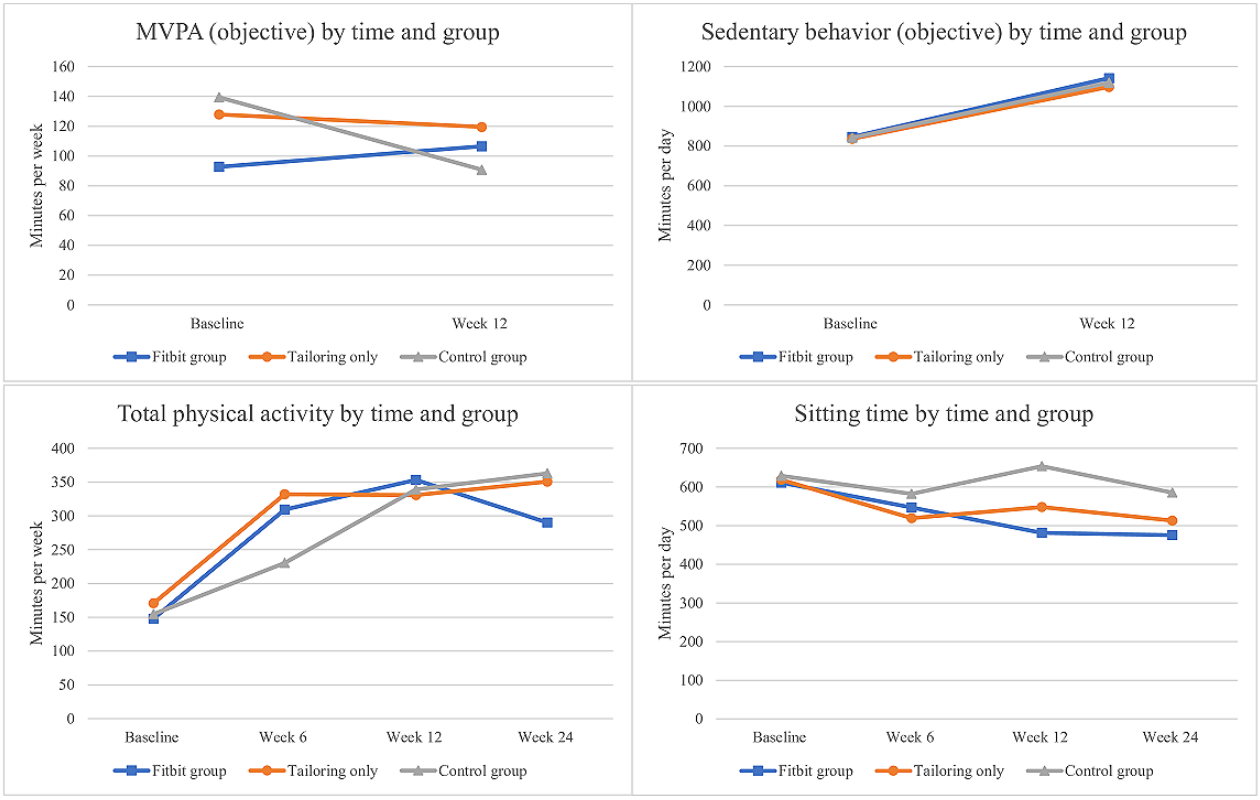
Table 2 and Figure 3 present the descriptives for accelerometer-measured MVPA and sedentary behavior and self-reported physical activity and sitting time by time and group. MVPA slightly increased in the tailoring+Fitbit group and decreased in the tailoring-only and control groups, whereas sedentary behavior increased in all groups. Self-reported physical activity more than doubled in all groups, and sitting time decreased in the 2 intervention groups. Sitting time remained relatively constant in the control group.

Table 2. Descriptives of physical activity and sedentary behavior by group and time (N=243).

	Tailoring+Fitbit group, mean (SD)	Tailoring only, mean (SD)	Control group, mean (SD)
Accelerometer-measured moderate to vigorous physical activity minutes per week			
Baseline (n=209)	92.75 (84.28)	127.89 (111.30)	139.44 (200.20)
Week 12 (n=141)	106.54 (127.47)	119.56 (102.13)	90.86 (98.00)
Self-reported total physical activity minutes per week			
Baseline (n=243)	147.95 (152.69)	170.62 (253.26)	154.64 (214.89)
Week 6 (n=173)	309.31 (202.28)	331.85 (336.85)	230.40 (278.45)
Week 12 (n=166)	353.21 (309.00)	330.68 (256.45)	339.02 (354.46)
Week 24 (n=157)	290.19 (268.09)	350.67 (261.73)	362.89 (362.85)
Accelerometer-measured sedentary minutes per day			
Baseline (n=209)	846.70 (242.17)	836.91 (206.95)	842.29 (230.64)
Week 12 (n=141)	1142.53 (173.58)	1098.43 (142.60)	1119.93 (149.46)
Self-reported sitting minutes per day			
Baseline (n=240)	611.14 (236.47)	618.47 (240.64)	629.04 (232.46)
Week 6 (n=172)	546.70 (257.83)	519.33 (234.15)	582.11 (273.23)
Week 12 (n=165)	481.14 (227.27)	547.77 (260.35)	653.70 (247.49)
Week 24 (n=156)	475.27 (233.65)	513.05 (253.75)	585.30 (252.34)
Accelerometer wear time in minutes per day			
Baseline (n=209)	1199.70 (203.04)	1212.84 (197.44)	1218.25 (190.86)
Week 12 (n=141)	1440.00 (0.00)	1440.00 (0.00)	1440.00 (0.00)

XSL•FO
RenderX

Figure 3. Moderate to vigorous physical activity (MVPA), total physical activity, sedentary behavior, and sitting time by time and group.



Main Analyses and Sensitivity Analyses

Tables 3 and 4 show the changes in accelerometer-measured MVPA, self-reported total physical activity, accelerometer-measured sedentary behavior, and self-reported sitting time by time and group. Results from the main analyses as well as the sensitivity analyses are presented. There was an overall time by group interaction for MVPA, but this was just above the criterion for significance (P=.05). No significant improvements in accelerometer-measured MVPA were observed from baseline to after the intervention within any group; however, the control group participants significantly decreased their MVPA by 35%. A pairwise difference was observed between the tailoring+Fitbit group and the control group, where the tailoring+Fitbit group participants increased their MVPA between baseline and week 12 by 59% more than those in the control group. This effect remained in the sensitivity analysis; however, the magnitude was reduced (43% difference). An overall time by group interaction was observed for self-reported total physical activity. All groups increased their self-reported physical activity from baseline to week 6, week 12, and week 24, and pairwise group comparisons revealed that this increase was 61% greater in the tailoring+Fitbit group than in the control group at 6 weeks. The effect was smaller in the sensitivity analysis (34% difference) and did not meet the criterion for significance.

No overall time by group interaction was observed for accelerometer-measured sedentary behavior. The tailoring+Fitbit and control groups significantly increased their accelerometer-measured sedentary behavior from baseline to after the intervention (by 57 and 56 minutes per day, respectively). The magnitude of these effects was lower in the sensitivity analysis (by 34 and 38 minutes per day, respectively) and did not meet the criterion for significance in the tailoring+Fitbit group. No pairwise group differences were observed. No overall time by group interaction was observed for self-reported sitting time. The tailoring+Fitbit group participants decreased their self-reported sitting time between baseline and week 6, 12, and 24 (by 67, 119, and 123 minutes per day, respectively). The tailoring-only group participants decreased their self-reported sitting time between baseline and week 6 and 24 (by 99 and 106 minutes per day, respectively). These effects remained in the sensitivity analysis; however, the magnitudes were reduced. Pairwise group comparisons revealed that the tailoring+Fitbit group participants decreased their sitting time between baseline and after the intervention 133 minutes per day more than those in the control group. However, the magnitude of effect was lower in the sensitivity analysis (98-minute decrease per day) and did not meet the criterion for significance.

XSL•FO
RenderX

Table 3. Main analysis, comparison of physical activity by time and group.

	Baseline to 6 weeks, estimate (95% CI)	*P* value	Baseline to 12 weeks, estimate (95% CI)	*P* value	Baseline to 24 weeks, estimate (95% CI)	*P* value	Group×time *P* value
Accelerometer-measured moderate to vigorous physical activity per week[a,b]	__[c]	—	—	—	—	—	.05
Tailoring+Fitbit (n=78)	—	—	1.03 (0.76 to 1.40)	.83	—	—	—
Tailoring only (n=96)	—	—	0.96 (0.71 to 1.30)	.79	—	—	—
Control (n=69)	—	—	0.65 (0.48 to 0.89)	.006	—	—	—
Tailoring+Fitbit vs control	—	—	1.59 (1.06 to 2.38)	.02	—	—	—
Tailoring only vs control	—	—	1.48 (0.99 to 2.20)	.06	—	—	—
Tailoring+Fitbit vs tailoring only	—	—	1.08 (0.73 to 1.59)	.71	—	—	—
Self-reported total physical activity per week[a]	—	—	—	—	—	—	.02
Tailoring+Fitbit (n=78)	2.31 (1.83 to 2.93)	<.001	2.24 (1.80 to 2.77)	.001	1.83 (1.44 to 2.32)	<.001	—
Tailoring only (n=96)	2.11 (1.64 to 2.72)	<.001	2.19 (1.71 to 2.80)	<.001	2.28 (1.82 to 2.87)	<.001	—
Control (n=69)	1.44 (1.07 to 1.93)	.01	1.99 (1.53 to 2.59)	<.001	2.24 (1.61 to 3.12)	<.001	—
Tailoring+Fitbit vs control	1.61 (1.11 to 2.33)	.01	1.12 (0.80 to 1.58)	.52	0.82 (0.55 to 1.22)	.32	—
Tailoring only vs control	1.47 (1.00 to 2.16)	.05	1.10 (0.77 to 1.58)	.61	1.02 (0.68 to 1.52)	.92	—
Tailoring+Fitbit vs tailoring only	1.10 (0.78 to 1.55)	.59	1.02 (0.74 to 1.41)	.90	0.80 (0.58 to 1.11)	.18	—
Accelerometer-measured sedentary time per day[b,d]	—	—	—	—	—	—	.79
Tailoring+Fitbit (n=78)	—	—	57.50 (9.65 to 105.34)	.02	—	—	—
Tailoring only (n=96)	—	—	38.66 (–6.10 to 83.42)	.09	—	—	—
Control (n=69)	—	—	56.07 (6.09 to 106.04)	.03	—	—	—
Tailoring+Fitbit vs control	—	—	1.43 (–66.67 to 69.54)	.96	—	—	—
Tailoring only vs control	—	—	–17.41 (–81.02 to 46.20)	.59	—	—	—
Tailoring+Fitbit vs tailoring only	—	—	18.84 (–42.03 to 79.70)	.55	—	—	—
Self-reported sitting time per day[d]	—	—	—	—	—	—	—
Tailoring+Fitbit (n=78)	–67.47 (–129.93 to –5.02)	.03	–119.36 (–190.19 to –48.53)	.001	–123.04 (–201.38 to –44.70)	.002	.13
Tailoring only (n=96)	–99.41 (–169.81 to –29.00)	.005	–68.58 (–144.76 to 7.60)	.08	–106.13 (–183.01 to –29.24)	.006	—
Control (n=69)	–58.29 (–120.50 to 3.93)	.06	13.68 (–50.66 to 78.01)	.68	–37.44 (–105.7 to 30.83)	.29	—

XSL•FO
RenderX

	Baseline to 6 weeks, estimate (95% CI)	*P* value	Baseline to 12 weeks, estimate (95% CI)	*P* value	Baseline to 24 weeks, estimate (95% CI)	*P* value	Group×time *P* value
Tailoring+Fitbit vs control	–9.19 (–96.16 to 77.79)	.84	–133.04 (–228.45 to –37.63)	.007	–85.60 (–187.84 to 16.64)	.10	—
Tailoring only vs control	–41.12 (134.26 to 52.02)	.38	–82.25 (–181.59 to 17.08)	.11	–68.68 (–171.22 to 33.85)	.19	—
Tailoring+Fitbit vs tailoring only	31.93 (–61.69 to 125.56)	.50	–50.78 (–153.91 to 52.34)	.33	–16.91 (–125.93 to 92.10	.76	—

[a]Reported as percentage change.

[b]Analyses controlled for accelerometer wear time.

[c]Not available.

[d]Reported as mean difference.

XSL•FO
RenderX

Table 4. Sensitivity analysis, comparison of physical activity by time and group.

	Baseline to 6 weeks, estimate (95% CI)	*P* value	Baseline to 12 weeks, estimate (95% CI)	*P* value	Baseline to 24 weeks, estimate (95% CI)	*P* value
Accelerometer-measured moderate to vigorous physical activity per week[a,b]	__[c]	—	—	—	—	—
Tailoring+Fitbit (n=78)	—	—	1.14 (0.90 to 1.42)	.26	—	—
Tailoring only (n=96)	—	—	1.02 (1.01 to 1.06)	.83	—	—
Control (n=69)	—	—	0.79 (0.60 to 1.05)	.10	—	—
Tailoring+Fitbit vs control	—	—	1.43 (1.02 to 2.01)	.04	—	—
Tailoring only vs control	—	—	1.28 (1.23 to 1.35)	.12	—	—
Tailoring+Fitbit vs tailoring only	—	—	1.12 (1.20 to 1.49)	.47	—	—
Self-reported total physical activity per week[a]	—	—	—	—	—	—
Tailoring+Fitbit (n=78)	2.27 (1.79 to 2.89)	<.001	2.36 (1.90 to 2.92)	<.001	2.08 (1.63 to 2.66)	<.001
Tailoring only (n=96)	2.10 (1.62 to 2.72)	<.001	2.23 (1.72 to 2.86)	<.001	2.25 (1.77 to 2.86)	<.001
Control (n=69)	1.68 (1.27 to 2.25)	<.001	2.20 (1.70 to 2.89)	<.001	2.48 (1.84 to 3.35)	<.001
Tailoring+Fitbit vs control	1.34 (0.92 to 1.93)	.12	1.06 (1.31 to 1.49)	.71	0.84 (0.58 to 1.22)	.36
Tailoring only vs control	1.23 (0.84 to 1.82)	.28	1.00 (0.69 to 1.45)	.99	0.90 (0.62 to 1.34)	.61
Tailoring+Fitbit vs tailoring only	1.08 (0.97 to 1.17)	.66	1.06 (0.76 to 1.49)	.71	0.93 (0.67 to 1.28)	.65
Accelerometer-measured sedentary time per day[b,d]	—	—	—	—	—	—
Tailoring+Fitbit (n=78)	—	—	34.49 (–0.33 to 69.31)	.05	—	—
Tailoring only (n=96)	—	—	24.12 (–2.02 to 50.27)	.07	—	—
Control (n=69)	—	—	38.09 (4.81 to 71.36)	.02	—	—
Tailoring+Fitbit vs control	—	—	–3.60 (–52.47 to 45.27)	.88	—	—
Tailoring only vs control	—	—	–13.96 (–55.92 to 27.99)	.51	—	—
Tailoring+Fitbit vs tailoring only	—	—	10.36 (–32.88 to 53.61)	.64	—	—
Self-reported sitting time per day[d]	—	—	—	—	—	—
Tailoring+Fitbit (n=78)	–56.95 (–123.61 to –9.72)	.09	–90.39 (–165.56 to –15.22)	.02	–109.73 (–191.90 to –27.56)	.009
Tailoring only (n=96)	–82.83 (–151.46 to –14.21)	.02	–56.09 (–130.50 to 18.32)	.14	–91.71 (–171.12 to –12.31)	.02
Control (n=69)	–62.49 (–136.69 to 11.71)	.10	7.57 (–64.09 to 79.24)	.83	–55.75 (–136.57 to 25.07)	.18
Tailoring+Fitbit vs control	5.54 (–90.15 to 101.24)	.91	–97.97 (–200.78 to 4.85)	.06	–53.97 (–169.52 to 61.57)	.36
Tailoring only vs control	–20.34 (–116.74 to 76.06)	.68	–63.66 (–161.89 to 34.56)	.20	–35.96 (–148.64 to 76.72)	.53
Tailoring+Fitbit vs tailoring only	25.89 (–68.65 to 120.43)	.59	–34.30 (–135.56 to 66.96)	.51	–18.01 (–123.33 to 87.30)	.74

[a]Reported as percentage change.

[b]Analyses controlled for accelerometer wear time.

[c]Not available.

[d]Reported as mean difference.

XSL•FO
RenderX

Discussion

Principal Findings

The main aim of the study was to determine the effectiveness of a computer-tailored physical activity intervention with Fitbit integration compared with a tailoring-only group and a control group at increasing MVPA from before to after the intervention. The second aim was to determine the effectiveness of a computer-tailored physical activity intervention with Fitbit integration compared with a tailoring-only group and a control group at increasing self-reported physical activity from before the intervention to the midintervention point, after the intervention, and follow-up. The findings showed that there were no significant MVPA changes in the tailoring+Fitbit group or tailoring-only group, whereas there was a decrease in the MVPA of the control group. MVPA increased more in the tailoring+Fitbit group than in the control group. All groups reported increasing their self-reported physical activity, and this increase was greater in the tailoring+Fitbit group than in the control group at the midintervention point. Together, these findings support past studies that have demonstrated that face-to-face, telephone, SMS text messaging, and email physical activity interventions using activity trackers are effective in older adults compared with a control group [25-27]. Most of these past studies focused on self-monitoring of steps and walking, whereas this study provided tailored feedback based on activity tracker–measured light-, moderate-, and vigorous-intensity physical activity. However, it should be noted that the Fitbit device had 5 lights, each of them indicating an additional 2000 steps reached for the day, which may have also motivated this group to maintain their physical activity, independent of the computer-tailored advice. Furthermore, the control group participants had a higher level of MVPA at baseline and therefore had more room to decrease their MVPA. This may have contributed to the between-group difference observed between the tailoring+Fitbit group and the control group on MVPA changes at week 12. Overall, these findings add to the literature by indicating that computer-tailored advice based on Fitbit measurement of light, moderate, and vigorous physical activity is likely to lead to improved physical activity outcomes compared with a control group.

The effectiveness of the tailoring+Fitbit intervention compared with a control group may be further improved by increasing the frequency of the feedback provided. Larsen et al [25] conducted a systematic review on physical activity trackers for older adults and found that only the interventions providing daily feedback on activity tracker data were effective. Our intervention provided in-depth feedback and theory-based behavior change support that would not be feasible to deliver daily. However, daily self-monitoring feedback on minutes of light, moderate, and vigorous physical activity might be feasible, in addition to the biweekly computer-tailored advice. Previous interventions for older adults have successfully delivered basic daily feedback on steps from activity trackers through smartphone apps in graphical and written form [44,45].

This study found no significant difference in accelerometer-measured MVPA or self-reported physical activity in the tailoring+Fitbit group compared with the tailoring-only group. This is not consistent with the findings of Vandelanotte et al [22], who found tailored advice based on Fitbit data to be significantly more effective at increasing self-reported physical activity compared with a tailoring-only group in middle-aged adults. It is possible that older adults do not benefit as much as younger adults from tailored advice based on Fitbit data, stemming from their lower interest in accelerometer-based activity trackers [23]. Future generations of older adults may be more familiar and interested in accelerometer-based activity trackers [24].

This study did not observe a difference in self-reported physical activity or accelerometer-measured MVPA over time between the tailoring-only and control groups. Previous research has demonstrated the overall effectiveness of computer-tailored physical activity advice in middle-aged adults [16], but the evidence in older adults is mixed. It has been demonstrated by 1 study that computer-tailored physical activity advice is effective in adults aged ≥65 years [20]. However, a study conducted in older adults with chronic diseases [19] also found that computer-tailored physical activity advice was not effective at increasing objectively measured physical activity in adults aged ≥65 years. A possible reason for these differences is the time frame of the interventions. The effective tailored intervention of Van Dyck et al [20] was 5 weeks in duration, whereas the ineffective intervention of Volders et al [19] and our intervention were 4 months and 3 months, respectively, in duration. The longer time frame for the postintervention assessment of these studies may have made it harder to detect group by time effects because improvements in physical activity tend to decline over time [16]. Intervention strategies to improve maintenance of physical activity over time may improve the effectiveness of computer-tailored physical activity advice in older adults [46]. More randomized controlled trials in adults aged ≥65 years are needed to determine the effectiveness of computer-tailored physical activity advice in older adults. The social cognitive theory postulates that both social support and individual cognition (eg, self-efficacy, outcome expectancies, and intentions) are important drivers of behavior [33]. In line with this, social support is important for health behavior change in older adults [47]. The lack of social support delivered through computer-tailored physical activity programs may also partially explain their limited effectiveness in adults aged ≥65 years. Future computer-tailored interventions for older adults may need to include additional components such as advice based on activity trackers or social support components to improve effectiveness.

The increase in objectively measured sedentary behavior in the tailoring+Fitbit group is not in line with a meta-analysis that found no intervention effect on sedentary behavior outcomes for either interventions targeting physical activity alone or those targeting both physical activity and sedentary behavior [48]. However, together these findings support that interventions must focus on sedentary behavior change rather than physical activity to see improvements in sedentary behavior. Our intervention targeted physical activity; however, sedentary behavior was briefly discussed in 2 modules where participants were encouraged to reduce their sitting time if it was >8 hours

XSL•FO
RenderX

a day. It is unlikely that the intervention's focus on MVPA as well as strength, balance, and flexibility exercises caused an increase in sedentary activities because the control group participants equally increased their sedentary behavior. The discrepancy between the objectively measured and self-reported sedentary behavior outcomes may be due to the participants' social desirability bias [49]. The improvement in self-reported sitting time at 12 weeks for the tailoring+Fitbit group compared with the control group might be because the tailoring+Fitbit group participants were more conscious about reducing sitting time, given that their sedentary behavior was being tracked. Of note, this did not translate into improvements in accelerometer-measured sedentary behavior.

Strengths and Limitations

The strengths of this study include the 3-group randomized design to determine individual effects of a tailored web-based physical activity intervention with and without Fitbit integration compared with a wait-list control group. The study objectively assessed physical activity at baseline and after the intervention. Although attrition at week 12 was moderate (77/243, 31.7%), this is comparable to many other trials examining web-based interventions [50-52]. Although the monetary incentive for completing the research surveys is unlikely to have had a large impact, it is possible that it increased participation in the intervention itself, going by the greater engagement in the overall study. The face-to-face meetings with researchers before and after the intervention may have also increased engagement in the intervention or helped to remove barriers to participation (eg, syncing the Fitbit device to the website). Therefore, lower engagement or additional barriers to participation may arise if the intervention is administered without monetary incentives or face-to-face meetings. The lack of accelerometer-assessed MVPA data at week 24 for the main outcome measure is a limitation. As such, we do not know whether the significant difference between the tailoring+Fitbit group and the control group at 12 weeks would remain at 24 weeks. The self-reported physical activity outcomes at 24 weeks suggest that physical activity changes were maintained in all groups, but this needs to be interpreted with caution because of the large differences in objectively measured and self-reported physical activity at baseline and week 12. Participants who completed the week 12 outcomes had a lower BMI than those who dropped out. Therefore, the findings cannot be generalized to older adults with a higher BMI. Another limitation is the lack of a Fitbit-only group with participants who receive a Fitbit device to track their physical activity without also receiving any tailored advice. This would help to determine whether the improvements in the tailoring+Fitbit group were due to being tracked by the Fitbit device or the combination of the tailored advice based on the Fitbit data. The Fitbit device had 5 lights, each of them indicating an additional 2000 steps reached for the day, which may have also motivated this group to maintain their activity, independent of the computer-tailored advice. Accelerometer wear time increased between baseline and week 12. Although analyses controlled for wear time, the increase in wear time may have had some effect on the decrease in MVPA and increase in sedentary behavior observed in some groups. Furthermore, the number of participants randomized to each group varied because of small numbers recruited within some randomization groups (eg, older men) with block sizes within each randomization group being 15. The control group participants had a higher level of MVPA at baseline and therefore had more room to decrease their MVPA. This may have contributed to the between-group difference observed between the tailoring+Fitbit and control groups on MVPA changes at week 12. Finally, the conservative a priori sample size calculation (n=300) was not met; however, we recruited 243 participants, which is comparable to similar studies [9], and there was enough power to detect MVPA group differences between the tailoring+Fitbit and control groups.

In conclusion, computer-tailored advice based on Fitbit measurement of physical activity in older adults is likely to lead to improved physical activity outcomes compared with no advice but not compared with advice based on self-reported physical activity. More research is needed to investigate ways to further improve effectiveness of computer-tailored advice based on Fitbit measurement in older adults.

Acknowledgments

SJA (102609) and SS (101240) are, and CV was (100427), supported by a fellowship from the National Heart Foundation of Australia. SS (1125586) and MJD (1141606) are supported by a fellowship from the National Health and Medical Research Council. The project was partially supported by a project grant from Central Queensland University and CV's fellowship from the National Heart Foundation of Australia (100427). These funding agencies had no role in the design of the study; collection, analysis, and interpretation of data; or in writing and submitting the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 809 KB - jmir_v24i5e31352_app1.pdf]

References

XSL•FO
RenderX

1. Bauman A, Merom D, Bull FC, Buchner DM, Fiatarone Singh MA. Updating the evidence for physical activity: summative reviews of the epidemiological evidence, prevalence, and interventions to promote "active aging". Gerontologist 2016 Apr;56 Suppl 2:S268-S280. [doi: 10.1093/geront/gnw031] [Medline: 26994266]
2. Hupin D, Roche F, Gremeaux V, Chatard JC, Oriol M, Gaspoz JM, et al. Even a low-dose of moderate-to-vigorous physical activity reduces mortality by 22% in adults aged ≥60 years: a systematic review and meta-analysis. Br J Sports Med 2015 Oct;49(19):1262-1267. [doi: 10.1136/bjsports-2014-094306] [Medline: 26238869]
3. National Health Survey: first results, 2017-2018. Australian Bureau of Statistics. 2019. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/national-health-survey-first-results/latest-release [accessed 2017-09-09]
4. The cost of physical inactivity. Medibank Private. 2007 Aug. URL: https://www.medibank.com.au/content/dam/client/documents/pdfs/pyhsical_inactivity.pdf [accessed 2021-12-01]
5. Productivity Commission. An ageing Australia: preparing for the future. Commonwealth of Australia. 2013 Nov. URL: https://www.pc.gov.au/research/completed/ageing-australia/ageing-australia.pdf [accessed 2017-11-24]
6. Older Australia at a glance. Cat. no. AGE 87. Australian Institute of Health and Welfare. 2018 Sep 10. URL: https://www.aihw.gov.au/getmedia/2cb104f4-c6d1-4728-9be3-a418840588de/Older-Australia-at-a-glance.pdf.aspx?inline=true [accessed 2021-12-01]
7. Murray JM, Brennan SF, French DP, Patterson CC, Kee F, Hunter RF. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: a systematic review and meta-analysis. Soc Sci Med 2017 Nov;192:125-133. [doi: 10.1016/j.socscimed.2017.09.021] [Medline: 28965003]
8. Digital lives of older Australians. Australian Communications and Media Authority. 2016. URL: http://www.acma.gov.au/theACMA/engage-blogs/engage-blogs/Research-snapshots/Digital-lives-of-older-Australians [accessed 2017-09-09]
9. Stockwell S, Schofield P, Fisher A, Firth J, Jackson SE, Stubbs B, et al. Digital behavior change interventions to promote physical activity and/or reduce sedentary behavior in older adults: a systematic review and meta-analysis. Exp Gerontol 2019 Jun;120:68-87. [doi: 10.1016/j.exger.2019.02.020] [Medline: 30836130]
10. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
11. Muellmann S, Forberger S, Möllers T, Bröring E, Zeeb H, Pischke CR. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review. Prev Med 2018 Mar;108:93-110. [doi: 10.1016/j.ypmed.2017.12.026] [Medline: 29289643]
12. Van Dyck D, Plaete J, Cardon G, Crombez G, De Bourdeaudhuij I. Effectiveness of the self-regulation eHealth intervention 'MyPlan1.0.' on physical activity levels of recently retired Belgian adults: a randomized controlled trial. Health Educ Res 2016 Oct;31(5):653-664. [doi: 10.1093/her/cyw036] [Medline: 27422898]
13. Broekhuizen K, de Gelder J, Wijsman CA, Wijsman LW, Westendorp RG, Verhagen E, et al. An Internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016 Apr 27;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]
14. Peels DA, van Stralen MM, Bolman C, Golsteijn RH, de Vries H, Mudde AN, et al. The differentiated effectiveness of a printed versus a web-based tailored physical activity intervention among adults aged over 50. Health Educ Res 2014 Oct;29(5):870-882. [doi: 10.1093/her/cyu039] [Medline: 24980023]
15. Rowley TW, Lenz EK, Swartz AM, Miller NE, Maeda H, Strath SJ. Efficacy of an individually tailored, internet-mediated physical activity intervention in older adults: a randomized controlled trial. J Appl Gerontol 2019 Jul;38(7):1011-1022. [doi: 10.1177/0733464817735396] [Medline: 29165018]
16. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
17. Alley SJ, Samra P, Rebar AL, Schoeppe S, Parkinson L, Power D, et al. A focus group study of older adults' perceptions and preferences towards web-based physical activity interventions. Inform Health Soc Care 2020 Sep;45(3):273-281. [doi: 10.1080/17538157.2019.1656210] [Medline: 31690152]
18. Ammann R, Vandelanotte C, de Vries H, Mummery WK. Can a website-delivered computer-tailored physical activity intervention be acceptable, usable, and effective for older people? Health Educ Behav 2013 Apr;40(2):160-170. [doi: 10.1177/1090198112461791] [Medline: 23077157]
19. Volders E, Bolman CA, de Groot RH, Verboon P, Lechner L. The effect of Active Plus, a computer-tailored physical activity intervention, on the physical activity of older adults with chronic illness(es)-a cluster randomized controlled trial. Int J Environ Res Public Health 2020 Apr 10;17(7):2590 [FREE Full text] [doi: 10.3390/ijerph17072590] [Medline: 32290072]
20. Van Dyck D, Herman K, Poppe L, Crombez G, De Bourdeaudhuij I, Gheysen F. Results of MyPlan 2.0 on physical activity in older Belgian adults: randomized controlled trial. J Med Internet Res 2019 Oct 07;21(10):e13219 [FREE Full text] [doi: 10.2196/13219] [Medline: 31593541]
21. Heesch KC, Hill RL, van Uffelen JG, Brown WJ. Are Active Australia physical activity questions valid for older adults? J Sci Med Sport 2011 May;14(3):233-237. [doi: 10.1016/j.jsams.2010.11.004] [Medline: 21276752]

XSL•FO
RenderX

22. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
23. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 07;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
24. Patel MS, Foschini L, Kurtzman GW, Zhu J, Wang W, Rareshide CA, et al. Using wearable devices and smartphones to track physical activity: initial activation, sustained use, and step counts across sociodemographic characteristics in a national sample. Ann Intern Med 2017 Nov 21;167(10):755-757. [doi: 10.7326/M17-1495] [Medline: 28973116]
25. Larsen RT, Christensen J, Juhl CB, Andersen HB, Langberg H. Physical activity monitors to enhance amount of physical activity in older adults - a systematic review and meta-analysis. Eur Rev Aging Phys Act 2019 May 4;16:7 [FREE Full text] [doi: 10.1186/s11556-019-0213-6] [Medline: 31073341]
26. Muellmann S, Buck C, Voelcker-Rehage C, Bragina I, Lippke S, Meyer J, et al. Effects of two web-based interventions promoting physical activity among older adults compared to a delayed intervention control group in Northwestern Germany: results of the PROMOTE community-based intervention trial. Prev Med Rep 2019 Sep;15:100958 [FREE Full text] [doi: 10.1016/j.pmedr.2019.100958] [Medline: 31410347]
27. S Oliveira J, Sherrington C, Zheng ERY, Franco MR, Tiedemann A. Effect of interventions using physical activity trackers on physical activity in people aged 60 years and over: a systematic review and meta-analysis. Br J Sports Med 2020 Oct;54(20):1188-1194. [doi: 10.1136/bjsports-2018-100324] [Medline: 31399430]
28. Alley S, van Uffelen JG, Schoeppe S, Parkinson L, Hunt S, Power D, et al. Efficacy of a computer-tailored web-based physical activity intervention using Fitbits for older adults: a randomised controlled trial protocol. BMJ Open 2019 Dec 23;9(12):e033305 [FREE Full text] [doi: 10.1136/bmjopen-2019-033305] [Medline: 31874890]
29. Cardinal BJ, Esters J, Cardinal MK. Evaluation of the revised physical activity readiness questionnaire in older adults. Med Sci Sports Exerc 1996 Apr;28(4):468-472. [doi: 10.1097/00005768-199604000-00011] [Medline: 8778552]
30. Physical activity and exercise guidelines for all Australians for older Australians (65 years and older). Department of Health, Australian Government. 2019. URL: https://www.health.gov.au/health-topics/physical-activity-and-exercise/physical-activity-and-exercise-guidelines-for-all-australians/for-older-australians-65-years-and-over [accessed 2020-12-16]
31. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9:52 [FREE Full text] [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
32. Ajzen I. Attitudes, personality, and behaviour. Homewood, IL, USA: The Dorsey Press; 2005.
33. Bandura A. Social foundations of thought and action: a social cognitive theory. Hoboken, NJ, USA: Prentice-Hall; 1986.
34. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
35. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, IMAGE Study Group. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011 Feb 18;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
36. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011 Feb;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
37. Kamada M, Shiroma EJ, Harris TB, Lee IM. Comparison of physical activity assessed using hip- and wrist-worn accelerometers. Gait Posture 2016 Feb;44:23-28 [FREE Full text] [doi: 10.1016/j.gaitpost.2015.11.005] [Medline: 27004628]
38. Tudor-Locke C, Barreira TV, Schuna Jr JM, Mire EF, Katzmarzyk PT. Fully automated waist-worn accelerometer algorithm for detecting children's sleep-period time separate from 24-h physical activity or sedentary behaviors. Appl Physiol Nutr Metab 2014 Jan;39(1):53-57. [doi: 10.1139/apnm-2013-0173] [Medline: 24383507]
39. Australian Institute of Health and Welfare. The Active Australia Survey: a guide and manual for implementation, analysis and reporting. Australian Governement. 2003 Apr 24. URL: https://www.aihw.gov.au/reports/physical-activity/active-australia-survey/contents/table-of-contents [accessed 2020-12-16]
40. Brown WJ, Trost SG, Bauman A, Mummery K, Owen N. Test-retest reliability of four physical activity measures used in population surveys. J Sci Med Sport 2004 Jun;7(2):205-215. [doi: 10.1016/s1440-2440(04)80010-0] [Medline: 15362316]
41. Bassett JK, Swain CT, Hodge AM, Mahmood S, Csizmadi I, Owen N, et al. Calibration of the Active Australia questionnaire and application to a logistic regression model. J Sci Med Sport 2021 May;24(5):474-480. [doi: 10.1016/j.jsams.2020.11.007] [Medline: 33281094]
42. Chau JY, van der Ploeg HP, Dunn S, Kurko J, Bauman AE. A tool for measuring workers' sitting time by domain: the Workforce Sitting Questionnaire. Br J Sports Med 2011 Dec;45(15):1216-1222. [doi: 10.1136/bjsports-2011-090214] [Medline: 21947817]
43. Eastin MS, LaRose R. Internet self efficacy and the psychology of the digital divide. J Comput Mediat Commun 2000 Sep;6(1):j.1083-6101.2000.tb00110.x. [doi: 10.1111/j.1083-6101.2000.tb00110.x]

XSL•FO
RenderX

44. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with chronic obstructive pulmonary disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
45. Skrepnik N, Spitzer A, Altman R, Hoekstra J, Stewart J, Toselli R. Assessing the impact of a novel smartphone application compared with standard follow-up on mobility of patients with knee osteoarthritis following treatment with Hylan G-F 20: a randomized controlled trial. JMIR Mhealth Uhealth 2017 May 09;5(5):e64 [FREE Full text] [doi: 10.2196/mhealth.7179] [Medline: 28487266]
46. Howlett N, Trivedi D, Troop NA, Chater AM. Are physical activity interventions for healthy inactive adults effective in promoting behavior change and maintenance, and which behavior change techniques are effective? A systematic review and meta-analysis. Transl Behav Med 2019 Jan 01;9(1):147-157 [FREE Full text] [doi: 10.1093/tbm/iby010] [Medline: 29506209]
47. Lindsay Smith G, Banting L, Eime R, O'Sullivan G, van Uffelen JG. The association between social support and physical activity in older adults: a systematic review. Int J Behav Nutr Phys Act 2017 Apr 27;14(1):56 [FREE Full text] [doi: 10.1186/s12966-017-0509-8] [Medline: 28449673]
48. Martin A, Fitzsimons C, Jepson R, Saunders DH, van der Ploeg HP, Teixeira PJ, EuroFIT consortium. Interventions with potential to reduce sedentary time in adults: systematic review and meta-analysis. Br J Sports Med 2015 Aug;49(16):1056-1063. [doi: 10.1136/bjsports-2014-094524] [Medline: 25907181]
49. Brenner PS, DeLamater JD. Social desirability bias in self-reports of physical activity: is an exercise identity the culprit? Soc Indic Res 2013 Jun 1;117(2):489-504. [doi: 10.1007/s11205-013-0359-y]
50. Kelders SM, Van Gemert-Pijnen JE, Werkman A, Nijland N, Seydel ER. Effectiveness of a web-based intervention aimed at healthy dietary and physical activity behavior: a randomized controlled trial about users and usage. J Med Internet Res 2011 Apr 14;13(2):e32 [FREE Full text] [doi: 10.2196/jmir.1624] [Medline: 21493191]
51. Skår S, Sniehotta FF, Molloy GJ, Prestwich A, Araújo-Soares V. Do brief online planning interventions increase physical activity amongst university students? A randomised controlled trial. Psychol Health 2011 Apr;26(4):399-417. [doi: 10.1080/08870440903456877] [Medline: 20830646]
52. 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] [Medline: 15829473]

Abbreviations

MVPA: moderate to vigorous physical activity

Edited by R Kukafka; submitted 18.06.21; peer-reviewed by J Boekhout-Oppeneer, S Bhattacharjya; comments to author 26.08.21; revised version received 14.01.22; accepted 08.02.22; published 12.05.22.

Please cite as:
Alley SJ, van Uffelen J, Schoeppe S, Parkinson L, Hunt S, Power D, Waterman N, Waterman C, To QG, Duncan MJ, Schneiders A, Vandelanotte C
The Effectiveness of a Computer-Tailored Web-Based Physical Activity Intervention Using Fitbit Activity Trackers in Older Adults (Active for Life): Randomized Controlled Trial
J Med Internet Res 2022;24(5):e31352
URL: https://www.jmir.org/2022/5/e31352
doi:10.2196/31352
PMID:35552166

©Stephanie J Alley, Jannique van Uffelen, Stephanie Schoeppe, Lynne Parkinson, Susan Hunt, Deborah Power, Natasha Waterman, Courtney Waterman, Quyen G To, Mitch J Duncan, Anthony Schneiders, Corneel Vandelanotte. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 12.05.2022. 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.

XSL•FO
RenderX

Original Paper

The Accuracy of Artificial Intelligence in the Endoscopic Diagnosis of Early Gastric Cancer: Pooled Analysis Study

Pei-Chin Chen[1,2*], MD, MSc; Yun-Ru Lu[2,3*], MD; Yi-No Kang[4,5,6,7*], MA; Chun-Chao Chang[8], MD

[1]Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[2]Department of General Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[3]Department of Anesthesiology, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[4]Evidence-Based Medicine Center, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

[5]Institute of Health Behaviors and Community Sciences, College of Public Health, National Taiwan University, Taipei, Taiwan

[6]Cochrane Taiwan, Taipei Medical University, Taipei, Taiwan

[7]Department of Health Care Management, College of Health Technology, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

[8]Division of Gastroenterology and Hepatology, Department of Internal Medicine, Taipei Medical University Hospital, Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chun-Chao Chang, MD
Division of Gastroenterology and Hepatology
Department of Internal Medicine
Taipei Medical University Hospital
No 252, Wuxing St
Taipei, 110
Taiwan
Phone: 886 227372181
Email: chunchao@tmu.edu.tw

Abstract

Background: Artificial intelligence (AI) for gastric cancer diagnosis has been discussed in recent years. The role of AI in early gastric cancer is more important than in advanced gastric cancer since early gastric cancer is not easily identified in clinical practice. However, to our knowledge, past syntheses appear to have limited focus on the populations with early gastric cancer.

Objective: The purpose of this study is to evaluate the diagnostic accuracy of AI in the diagnosis of early gastric cancer from endoscopic images.

Methods: We conducted a systematic review from database inception to June 2020 of all studies assessing the performance of AI in the endoscopic diagnosis of early gastric cancer. Studies not concerning early gastric cancer were excluded. The outcome of interest was the diagnostic accuracy (comprising sensitivity, specificity, and accuracy) of AI systems. Study quality was assessed on the basis of the revised Quality Assessment of Diagnostic Accuracy Studies. Meta-analysis was primarily based on a bivariate mixed-effects model. A summary receiver operating curve and a hierarchical summary receiver operating curve were constructed, and the area under the curve was computed.

Results: We analyzed 12 retrospective case control studies (n=11,685) in which AI identified early gastric cancer from endoscopic images. The pooled sensitivity and specificity of AI for early gastric cancer diagnosis were 0.86 (95% CI 0.75-0.92) and 0.90 (95% CI 0.84-0.93), respectively. The area under the curve was 0.94. Sensitivity analysis of studies using support vector machines and narrow-band imaging demonstrated more consistent results.

Conclusions: For early gastric cancer, to our knowledge, this was the first synthesis study on the use of endoscopic images in AI in diagnosis. AI may support the diagnosis of early gastric cancer. However, the collocation of imaging techniques and optimal algorithms remain unclear. Competing models of AI for the diagnosis of early gastric cancer are worthy of future investigation.

Trial Registration: PROSPERO CRD42020193223; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=193223

(J Med Internet Res 2022;24(5):e27694) doi:10.2196/27694

XSL•FO
RenderX

KEYWORDS

artificial intelligence; early gastric cancer; endoscopy

Introduction

Gastric cancer is the fifth most common cancer and the third leading cause of cancer deaths worldwide, contributing to 19.1 million disability-adjusted life years in 2017 [1,2]. Its primary risk factors are *Helicobacter pylori* infection and a family history of gastric cancer [3,4]. Despite advancements in endoscopic, surgical, and systemic therapies, the global 5-year survival rate of those with gastric cancer remains low (25%-30%) [5]. Gastric cancer has an excellent prognosis at early stages, with a 5-year survival rate of approximately 95%, but it has a median survival rate of less than one year at advanced stages [6,7]. Its favorable early prognosis is reflected in the lower mortality rates of gastric cancer in East Asia, which can be ascribed to the implementation of nationwide screening [8]. This reinforces the importance of early diagnosis. However, gastrointestinal endoscopy, the standard detection method for early gastric cancer, has an unsatisfactory sensitivity of 70% and is operator dependent [9]. Despite efforts to increase the detection rate, a valid screening method has yet to be developed [10,11]. The recent advancement in artificial intelligence (AI) systems, which provides highly accurate and efficient image recognition, may indicate a solution to this problem.

Although significant increases in AI exist in many fields and in health care [12-19], AI has various definitions [20]. According to the cognitive modeling approach, AI can be seen as machines that perform or exhibit actions corresponding to intelligence such as human behavior [20,21]. Machine learning, a subset of AI, involves studying how computers learn to improve task performance through experience without being programmed. This learning is achieved through various approaches. For instance, support vector machines, widely used in data classification, are machine learning algorithms that work by calculating the best separating plane for distinguishing between different objects. Deep learning, another machine learning method, simulates the multiple hierarchical layers of neural networks to make decisions based on features extracted from massive training data. Convolutional neural networks are deep learning algorithms primarily used in image recognition [22].

Since the breakthrough of deep learning in the 2010s, the use of AI in clinical practice has increased dramatically [22,23], and many studies have applied AI for screening or diagnosis [24-27]. Several studies have provided promising results for the AI-assisted endoscopic diagnosis of gastric cancer [28]. In a multicenter case control study of 84,424 participants, a deep learning–aided system demonstrated a detection rate of upper gastrointestinal cancer comparable to that of an expert endoscopist [29]. Other studies have investigated the diagnostic accuracy of AI for gastric polyps and the invasion depth of gastric cancers [30,31]. Nevertheless, the rate of detection of early gastric cancer, which allows for prompt intervention and increased survival rates, remains low. Multiple studies on the AI-assisted diagnosis of early gastric cancer have been conducted in the past 5 years, but results have been inconsistent and highly variable. Furthermore, the role of AI in early gastric cancer is more important than in advanced gastric cancer since early gastric cancer is not easily identified in clinical practice; however, to our knowledge, past syntheses appear to have limited focus on the population with early gastric cancer. Thus, we investigated the performance of AI-assisted endoscopic diagnosis of early gastric cancer.

Methods

Definition

Early gastric cancer was defined as mucosal and submucosal (T1) gastric cancer irrespective of lymph node involvement. Studies involving advanced gastric cancer, precancerous lesions such as intestinal metaplasia and dysplasia, and gastric cancer without specific annotations were excluded. The accuracy of AI was defined as the area under the hierarchical summary receiver operating characteristic curve or the area under the curve (AUC).

Study Search and Selection Strategy

This meta-analysis was performed according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We systematically searched the PubMed, Embase, Cochrane Library, and Web of Science databases for studies that assessed the diagnostic accuracy of AI in early gastric cancer from endoscopic images from database inception to June 2020. We used “gastric cancer,” “endoscopy,” and “artificial intelligence” as relevant terms with Boolean operators “OR” and “AND” (Multimedia Appendix 1). Two authors, P-CC and L-YR, independently screened the study titles and abstracts. Studies that used AI to diagnose early gastric cancer from endoscopic images were included. Studies that did not provide a 2×2 contingency table were not included in the final analysis. This study was registered in PROSPERO (registration CRD42020193223).

Study Quality Assessment and Data Extraction

The quality of the included studies was assessed independently by 2 authors (P-CC and L-YR) on the basis of the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2), and all disagreement was resolved through discussion with the third author (Y-NK). The assessment included risk of bias and applicability to the QUADAS-2 domains: patient selection, index test, reference standard, and flow and timing. From the included studies, we extracted data on the number of endoscopic images of lesions diagnosed as early gastric cancer (ie, true positive), the number of endoscopic images of benign lesions misdiagnosed as malignant (ie, false positive), the number of endoscopic images of malignant lesions misdiagnosed as benign (ie, false negative), and the number of endoscopic images of benign lesions correctly diagnosed as benign (ie, true negative). We also extracted data on the country of origin, AI methods, and image modalities used.

XSL•FO
RenderX

Study Outcomes and Statistical Analysis

The primary outcome was the accuracy of AI to diagnose early gastric cancer from endoscopic images. Secondary outcomes focused on the sensitivity analysis of (a) different AI methods, (b) endoscopic imaging modalities, (c) studies that compared AI and endoscopist performance, (d) studies that evaluated larger gastric lesions (>20 mm), (e) studies that simply differentiated abnormal and normal lesions rather than using pathological staging, and (f) studies that separated the training and testing data sets during AI training. Sensitivity analysis was conducted if a subgroup contained more than two studies. We only assessed the heterogeneity of the included studies. Following extraction, the data were primarily analyzed using STATA 14 (StataCorp LP, StataCorp) except for subgroups with fewer than four studies. The midas and metandi commands were used to determine sensitivity, specificity, and AUC and analyze the summary receiver operating characteristic (SROC) and hierarchical summary receiver operating characteristic (HSROC) curves. Basic formulas for the analyses were as follows:

$$\ln DOR = (\text{logit } TPR) - (\text{logit } FPR) \quad (1)$$

$$\text{proxy for the threshold} = (\text{logit } TPR) - (\text{logit } FPR) \quad (2)$$

$$TPR \text{ of } SROC = 1/[1/(1+ea/(1-b)) \times (FPR/(1-FPR))(1+b)/(1-b)] \quad (3)$$

In the formulas, “a” is the intercept, “b” is slope, and DOR refers to the diagnostic odds ratio. Moreover, TPR is the true positive rate, and FPR is the false-positive rate. The modchk tool was used to examine goodness-of-fit and bivariate normality before SROC analysis in a bivariate mixed-effects model. The metabias command and the pubbias syntax were used to perform the Egger test and Deeks funnel plot asymmetry tests, respectively. The Egger test for diagnostic meta-analysis was based on the formula proposed by Hasselblad and Hedges, and the formula is mainly to detect publication bias detection via testing standard normal deviate among the included studies [32,33].

$$\text{standard normal deviate} = a + b \times SE_{(d)}^{-1} \quad (4)$$

In the regression model, with intercept “*a*” and slope “*b*,” the standard normal deviation could be estimated by using diagnostic *d* divided by SE of the diagnostic *d*. The metaprop package in STATA was mainly used to synthesize the sensitivity and specificity. I^2 statistics were used to determine levels of heterogeneity via the formula as follows:

$$I2 = ((Q - df)/Q) \times 100 \quad (5)$$

where Q refers to Cochran Q, and *df* is the degree of freedom. Because R software (The R Foundation) does not restrict the number of observations used in the meta-analysis, it was used for sensitivity analysis if subgroups consisted of fewer than four studies. Indeed, a meta-analysis in R could be carried out when more than two studies report the same outcome by pooling data with logit transformation and Clopper-Pearson interval method (also called exact binomial interval) based on inverse variance. Function metaprop in package meta for R was applied to carry out sensitivity analysis, and the mada package in R was used to calculate the pooled accuracy. Besides, the metagen package in R was used to synthesize endoscopist performance because of the lack of detailed data on each endoscopist.

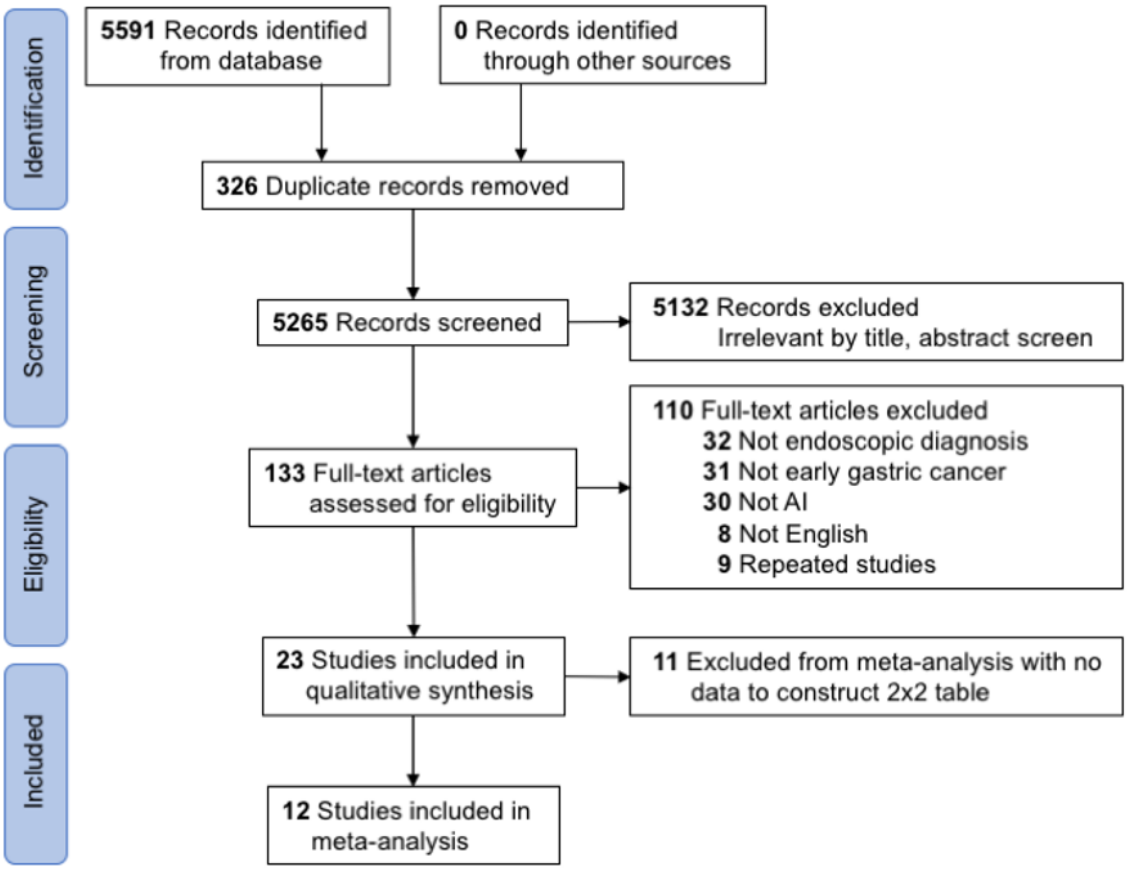
Results

Literature Search and Review

Of the 5591 studies identified in the literature review, 5265 underwent title and abstract screening after duplication removal. The flowchart of the literature review process was constructed according to the PRISMA flowchart format (Figure 1). We excluded 5132 irrelevant studies and assessed the eligibility of the remaining 133 studies through full-text reading. Studies evaluating nonearly gastric cancer (eg, advanced gastric cancer and metaplasia) were excluded. Overall, 23 studies investigated the performance of AI on early gastric cancer diagnosis from endoscopic images. Finally, 12 studies comprising a total of 11,685 cases were included in the meta-analysis [34-45].

XSL•FO
RenderX

Figure 1. Flowchart of the study selection process according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) format. AI: artificial intelligence.



Study Description and Bias Assessment

Detailed information on the 12 studies is listed in Table 1. All studies were conducted in Asia, including Japan (k=8), China (k=2), and Korea (k=2), in or after 2012. All were case control studies with testing data sets containing 81 to 3390 images. Patients in 10 studies had pathological proof of early gastric cancer, whereas in the other 2 studies, the endoscopic images were collected through description. White light imaging (WLI), narrow-band imaging (NBI), flexible spectral imaging color enhancement, and mixed imaging modalities were used in 4 (33%), 2 (17), 1 (8%), and 2 (17%) studies, respectively. Moreover, 8 (67%) studies used deep learning methods (eg, convolutional neural networks) as their AI backbone, and 3 (25%) studies employed nondeep learning methods (support vector machines and discriminant analysis of principal components). Comparisons of the diagnostic performance of AI and endoscopists were conducted in 3 (25%) studies, and 2 (17%) studies included endoscopic images of small lesions (<20 mm) in early gastric cancer. In 3 (25%) studies, the training and testing data sets were not separated for AI training. Table 1 presents a detailed description of the 12 studies.

We also assessed the quality of the studies along with the risk of bias according to the revised QUADAS-2 tool (Multimedia Appendix 2). All studies, including the 3 that failed to separate the training and testing data sets, had high bias risks for patient selection because of their retrospective design. Moreover, 2 (17%) studies assessed early gastric cancer but did not mention pathological staging. Thus, they were classified as having a high risk of bias for the index test.

XSL•FO
RenderX

Table 1. Characteristics of the included studies.

Study ID	Country of origin	Testing image number	Reference standard	Image modality	AI[a] method	AI training and testing data set	Standard reference	Endoscopist comparison	Other information
Kubota et al, 2012 [43]	Japan	902	Pathology	Not mentioned[b]	Multilayer neural network	Not separated	Unclear	No	Detected with pathological grading prediction
Miyaki et al, 2013 [44]	Japan	92	Pathology	FICE[c]	SVM[d] (scale-invariant feature transform)	Separated	Pathology	No	Differentiated early gastric cancer from noncancerous tissues
Liu et al, 2016 [41]	China	400	Pathology	Not mentioned[b]	Principal component discriminant analysis (YCbCr color space)	Separated	Pathology	No	Differentiated early gastric cancer from normal tissues
Kanesaka et al, 2018 [37]	Japan	81	Pathology	NBI[e]	SVM (grey-level co-occurrence feature)	Separated	Pathology	No	Included only depressed type early gastric cancers that were <10 mm in size
Sakai et al, 2018 [36]	Japan	926	Pathology	WLI[f]	CNN[g] (GoogLeNet)	Not separated	Pathology	No	—[h]
Yamakawa et al, 2018 [45]	Japan	817	Unclear[i]	Not mentioned[j]	Not mentioned	Separated	Unclear	No	Differentiated early gastric cancer from nonneoplastic tissues
Cho et al, 2019 [35]	Korea	200	Pathology	WLI	CNN (Inception-Resnet-v2)	Separated	Pathology	Yes	Detected early gastric cancer with pathological grading prediction
Namikawa et al, 2019 [34]	Japan	1479[j]	Unclear[i]	WLI, NBI, Chromo[k]	CNN	Separated	Pathology	No	Differentiated early gastric cancer from gastric ulcers
Wu et al, 2019 [39]	China	200	Pathology	WLI, NBI, BLI[l]	CNN (VGG16 + Resnet-50)	Separated	Pathology	Yes	Differentiated early gastric cancer from gastritis and normal tissues
Yoon et al 2019 [42]	Korea	3390	Pathology	WLI	CNN (VGG16)	Not separated	Pathology	No	—
Horiuchi et al, 2020 [38]	Japan	258	Pathology	NBI	CNN (GoogLeNet)	Separated	Pathology	No	Differentiated early gastric cancer from *Helicobacterpylori*–related gastritis
Ikenoyama et al, 2020 [40]	Japan	2940	Pathology	WLI	CNN (Single-shot multiBox Detector)	Separated	Pathology	Yes	Included only early gastric lesions that were <20 mm

[a]AI: artificial intelligence.

[b]Studies that failed to mention imaging modalities.

[c]FICE: flexible spectral imaging color enhancement.

[d]SVM: support vector machine.

[e]NBI: narrow-band imaging.

[f]WLI: white light imaging.

[g]CNN: convolutional neural network.

[h]Not available.

[i]Studies that mentioned early gastric cancer but without reference to pathological staging.

XSL•FO
RenderX

[j]Studies were reported in meeting abstracts.

[k]Chromo: chromoendoscopy.

[l]BLI: blue laser imaging.

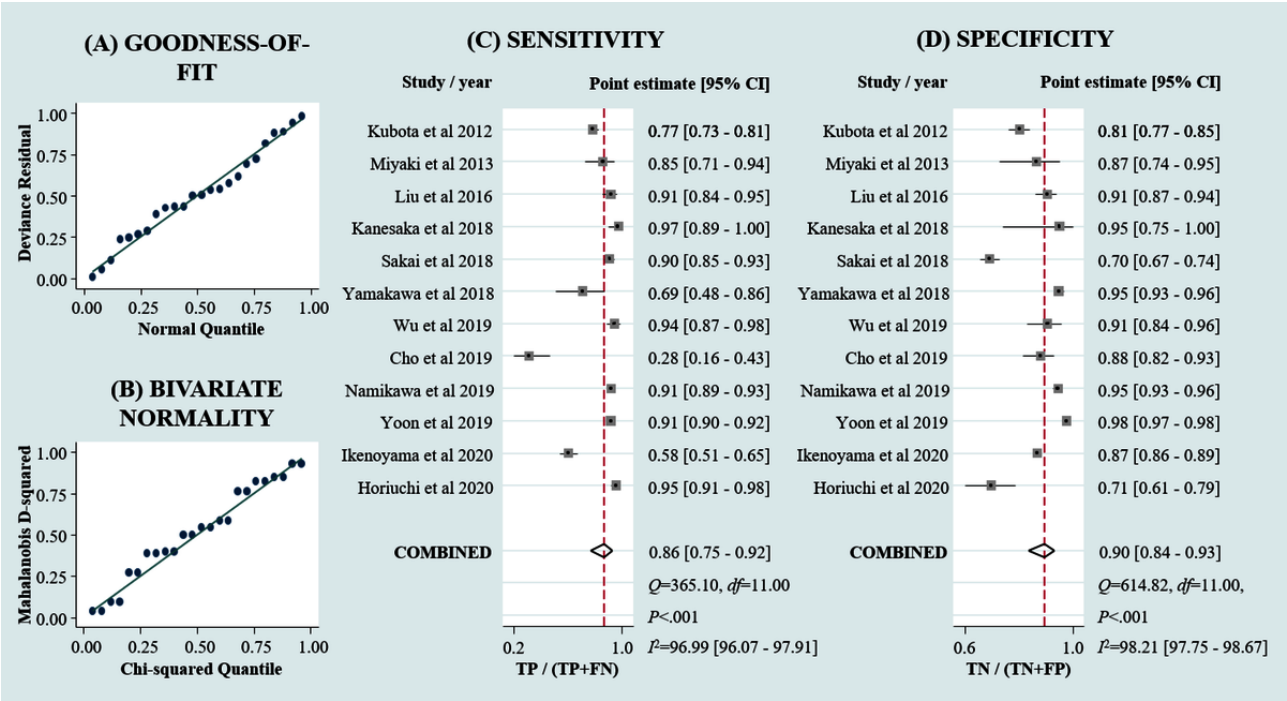
Diagnostic Performance of AI for Early Gastric Cancer

To assess the diagnostic ability of AI to detect early gastric cancer from endoscopic images, we performed a meta-analysis on the selected 12 studies. Goodness-of-fit (Figure 2A) and bivariate normality (Figure 2B) demonstrated that the included data were appropriate for further analysis. The pooled sensitivity and specificity of AI were 0.86 (95% CI 0.75-0.92) and 0.90 (95% CI 0.84-0.93), respectively (Figures 2C and 2D). Empirical Bayesian predictions were consistent with the observed sensitivity and specificity (Multimedia Appendix 3). Highly heterogeneous estimates (I^2>90%) necessitated subgroup analysis and sensitivity analysis. Laminated figures of the SROC and HSROC plots indicate an AUC of 0.94 (95% CI 0.92-0.96) with a confidence region (Figure 3A). However, the scatter matrix (Multimedia Appendix 4) suggests that in clinical practice, diagnosis of early gastric cancer may not substantially benefit from AI assistance. The Deeks funnel plot asymmetry test (Figure 3B) and Egger test (Multimedia Appendix 5) did not detect significant publication bias in the pooled results of AI-assisted diagnosis of early gastric cancer.

We assessed the diagnostic performance of various AI methods and endoscopic imaging modalities for early gastric cancer (Table 2). The pooled sensitivity and specificity in studies using deep learning methods were 0.84 (95% CI 0.69-0.93) and 0.88 (95% CI 0.80-0.93), respectively. Studies using nondeep learning methods had a pooled sensitivity and specificity of 0.91 (95% CI 0.86-0.95) and 0.90 (95% CI 0.87-0.93), respectively. The accuracy of the nondeep learning group (AUC=0.96) was higher than that of the deep learning group (AUC=0.93; Multimedia Appendices 6 and 7).

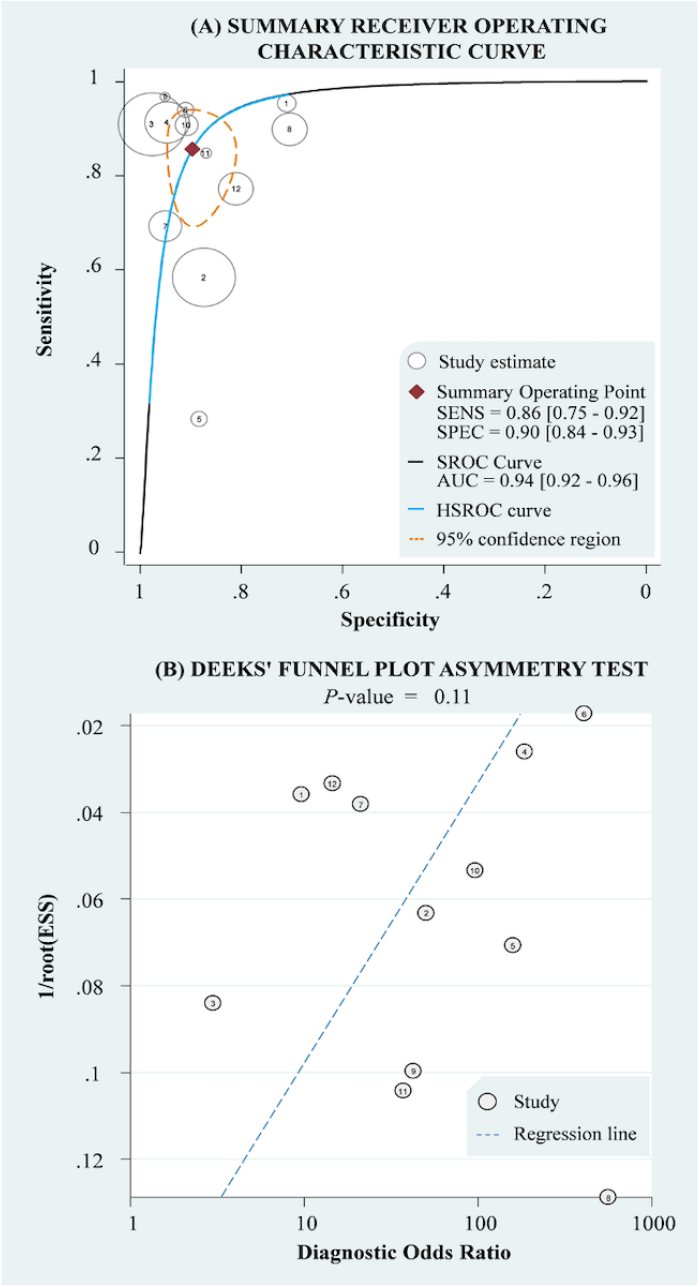
For endoscopic imaging modalities, studies using WLI had a sensitivity and specificity of 0.73 (95% CI 0.42-0.91) and 0.89 (95% CI 0.76-0.96), respectively. Studies using NBI reported a sensitivity and specificity of 0.96 (95% CI 0.92-0.98) and 0.83 (95% CI 0.54-0.95), respectively. The accuracy of the NBI group (AUC=0.96) was higher than that of the WLI group (AUC=0.90; Multimedia Appendices 8 and 9). Table S1 (Multimedia Appendix 10) shows a comparison of the diagnostic performance of AI and endoscopists for early gastric cancer from the three studies (n=91).

Figure 2. Overall sensitivity and specificity of artificial intelligence–assisted diagnosis of early gastric cancer. (A) Goodness-of-fit; (B) bivariate normality; (C) forest plot of overall sensitivity; and (D) forest plot of overall specificity. FP: false positive; TN: true negative.



XSL•FO
RenderX

Figure 3. Summary receiver operating characteristic curve, HSROC, AUC, and the Deeks funnel plot asymmetry test of artificial intelligence–assisted diagnosis of early gastric cancer. AUC: area under the curve; ESS: effective sample sizes; HSROC: hierarchical summary receiver operating characteristic; SENS: sensitivity; SPEC: specificity; SROC: summary receiver operator characteristic.



Additional Analysis

We excluded some studies with a high risk of bias and performed sensitivity analysis on the remaining studies (Tables S2-S5 Multimedia Appendices 11-14). Furthermore, we also examined how the results were affected by studies with unknown AI methods. Sensitivity analyses indicated that pooled estimates were not seriously affected by the factors (Table 2). Lower heterogeneity and specificity were observed in endoscopist performance when we excluded studies that only evaluated small lesions and studies that predicted pathological staging (Tables S2 and S3 in Multimedia Appendices 11 and 12). Lower heterogeneity was also noted in WLI subgroups if the training and testing data sets were separated for AI training (Table S4 in Multimedia Appendix 13). No other additional analyses provided credible evidence.

XSL•FO
RenderX

Table 2. Pooled sensitivity, specificity, and accuracy of the studies included in the meta-analysis and sensitivity analysis.

Group (studies and number of patients)	Sensitivity (95% CI)	I^2, %	Specificity (95% CI)	I^2, %	AUC[a]
Overall (12 studies, n=11,685)	0.86 (0.75-0.92)	97	0.90 (0.84-0.93)	97	0.94
Subgroup analysis on different AI[b] methods					
Deep learning (8 studies, n=10,295)	0.84 (0.69-0.93)	98	0.88 (0.80-0.93)	98	0.93
Nondeep learning (3 studies, n=573)	0.91 (0.86-0.95)	18	0.90 (0.87-0.93)	0	0.96
Subgroup analysis on various imaging modalities					
WLI[c] (4 studies, n=7456)	0.73 (0.42-0.91)	99	0.89 (0.76-0.96)	99	0.902
NBI[d] (2 studies, n=339)	0.96 (0.92-0.98)	0	0.83 (0.54-0.95)	51	0.959
Sensitivity analysis					
Excluding studies with unknown method (11 studies, n=10,868)	0.87 (0.76-0.93)	97	0.89 (0.83-0.93)	97	0.936
Excluding studies with sample size <100 (10 studies, n=11,512)	0.84 (0.71-0.92)	97	0.89 (0.83-0.94)	98	0.932
Excluding studies without separation of testing data (9 studies, n=6467)	0.85 (0.70-0.93)	96	0.90 (0.86-0.93)	91	0.934
Excluding studies with any situation abovementioned (6 studies, n=5477)	0.84 (0.62-0.94)	98	0.89 (0.83-0.93)	92	0.923

[a]AUC: area under the curve.

[b]AI: artificial intelligence.

[c]WLI: white light imaging.

[d]NBI: narrow-band imaging.

Discussion

Principal Findings

To our knowledge, this was the first systematic review and meta-analysis of AI-assisted endoscopic diagnosis of early gastric cancer. The accuracy, sensitivity, and specificity were 0.94, 0.86, and 0.90, respectively. High heterogeneity was noted. Sensitivity analysis revealed less heterogeneity in studies using nondeep learning AI methods and endoscopic NBI.

Our results indicate good sensitivity and specificity of AI-assisted detection of early gastric cancer. However, high heterogeneity was also noted among the included studies, which may be attributed to between-study differences in machine learning methods and imaging modalities [46]. In a meta-analysis of AI prediction of colonic polyp histology, AI performance was better when deep learning was used as a backbone and when NBI was used to identify the lesions [46]. In this study, we also investigated the roles of various machine learning methods and imaging modalities. Unfortunately, only 2 studies in the deep learning subgroup used the same deep learning algorithm, and no two studies in the nondeep learning subgroup classified the lesions according to the same features. Only 6 studies specified their endoscopic imaging modalities. Less heterogeneity was observed in the nondeep learning and NBI groups, possibly because of the compliance of early gastric cancer diagnosis to the vessel plus surface classification system under NBI. This indicates that nondeep learning methods and NBI may provide more consistent results and can be applied in clinical practice earlier than deep learning methods and WLI. Further investigations are warranted.

We assessed the diagnostic performance of AI and endoscopists (n=91) for early gastric cancer detection, which was compared in 3 studies. The endoscopists were assigned to only 1 subgroup because of the inconsistent definitions of expert and nonexpert endoscopists between studies. The sensitivity and specificity of AI were 0.67 and 0.87, respectively, and those of the endoscopists were 0.68 and 0.92, respectively. In both groups, diagnostic performance varied widely with high heterogeneity. The diagnostic performance of AI was better than that of WLI compared with other studies; a meta-analysis reported a pooled sensitivity and specificity of 48% and 67% between endoscopists and WLI, whereas those between endoscopists and NBI were 83% and 97%, respectively [47]. In this study, AI and endoscopist performance were comparable in individual studies, but this effect diminished when studies were pooled. Further research comparing AI and endoscopist performance for early gastric cancer diagnosis is required.

Only 2 of the included studies evaluated only small lesions [37,40]. Smaller lesions and mucosal lesions were less accurately detected by AI [42]. Kanesaka et al [37] included only depressed and small (<10 mm) lesions, and the AI system of nondeep learning methods was trained using a small data set of 126 images from NBI. In another study, early gastric cancer lesions less than 20 mm in diameter were included in the WLI testing data set, and the deep learning AI system was trained using a data set of 13,584 images of early and advanced gastric cancer [40]. Because these 2 studies used distinct materials and methods, their findings may not be representative. The accuracy of AI-assisted detection of small gastric cancer lesions warrants further investigation.

XSL•FO
RenderX

Some studies have explored the application of AI to other aspects of gastroendoscopy. For example, Wu et al [39] used AI to monitor endoscopic blind spots and identify regions indicative of early gastric cancer. A randomized controlled trial in China reported that AI reduced the rate of endoscopic blind spots [48]. Other studies have tested the accuracy of AI in predicting the invasion depth of gastric cancer—conventionally assessed through endoscopic ultrasound—from endoscopic images. In their study of AI-assisted simultaneous detection of gastric cancer and invasion depth, Yoon et al [42] reported a sensitivity and specificity of invasion depth of 79.2% and 77.8%, respectively. In a study by Zhu et al [31], the predicted sensitivity and specificity from the T1 to the T4 stage were 76% and 96%, respectively. Nevertheless, relevant evidence is limited, and further investigation is required.

The considerable advancement of AI in precise image recognition challenges the roles of physicians in disease diagnosis. AI systems offer certain advantages over physician diagnosis, the foremost of which are faster image processing rates and continuous work. In all included studies that specified image processing time, that of AI systems was shorter than that of endoscopists. AI assistance may reduce the risk of human error that arises from performing numerous endoscopic examinations. Moreover, the training of AI systems is considerably faster and less complicated than that of endoscopists. Well-trained AI systems learn from analyzing numerous images, whereas endoscopists rely on their individual skills and clinical experience. Training endoscopists is expensive and time-consuming because of the steep learning curve for the various image-enhancing techniques. In addition, AI may work as a double-check system during or after endoscopy, given its high sensitivity and specificity. AI allows for a second opinion, which is particularly valuable now that gastroendoscopy has been popularized and nationwide screening for gastric cancer has been implemented.

Limitations

Our study had several limitations. First, all the included studies were retrospective case control studies performed in Asia, some of which compared early gastric cancer and normal gastric tissues, and some compared benign gastric lesions such as ulcers and gastritis. The possibility of selection bias cannot be ruled out. A randomized controlled trial comparing the diagnostic performance of AI and endoscopists for early and advanced gastric cancer (NCT04040374) is currently underway. Second, all the studies identified gastric lesions from still, clear, endoscopic images; images with blood or mucus were excluded. In daily practice, however, gastroendoscopy is recorded in video format, and still images are only captured for suspicious lesions. Blood, food debris, mucus, and foam, which reduce the accuracy of AI, are commonly encountered during examination [39]. Several studies have reported excellent accuracy of AI systems in recognizing gastric cancer from endoscopic video [39,49]. However, further studies and faster image processing rates are necessary. Third, our pooled estimates were highly heterogeneous, and the subgroup and sensitivity analyses did not substantially reduce heterogeneity. The statistical heterogeneity may be ascribed to differences in the AI methods and endoscopic imaging techniques. These potential sources of heterogeneity should be discussed in future research. At present, AI may assist endoscopists in double-checking suspicious lesions.

Conclusions

To our knowledge, this is the first meta-analysis of the performance of AI in detecting early gastric cancer using endoscopic images. The available evidence suggests that AI can support the diagnosis of early gastric cancer; however, the collocation of imaging techniques and optimal algorithm remains unclear. Larger prospective cohort studies should be conducted to further validate the diagnostic accuracy of AI. Moreover, competing models of AI for the detection of early gastric cancer are worthy of future investigation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary File 1. Search strategy (primary search strategy).
[PDF File (Adobe PDF File), 402 KB - jmir_v24i5e27694_app1.pdf]

Multimedia Appendix 2
Supplementary File 2. Study quality assessment according to the QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies [revised]).
[PDF File (Adobe PDF File), 437 KB - jmir_v24i5e27694_app2.pdf]

Multimedia Appendix 3
Supplementary File 3. Forest plot of empirical Bayes predicted and observed findings.
[PDF File (Adobe PDF File), 564 KB - jmir_v24i5e27694_app3.pdf]

Multimedia Appendix 4
Supplementary File 4. Scatter matrix.

XSL•FO
RenderX

[PDF File (Adobe PDF File), 463 KB - jmir_v24i5e27694_app4.pdf]

Multimedia Appendix 5
Supplementary File 5. Egger's test.
[PDF File (Adobe PDF File), 452 KB - jmir_v24i5e27694_app5.pdf]

Multimedia Appendix 6
Supplementary File 6. Subgroup analysis for studies that used deep learning.
[PDF File (Adobe PDF File), 511 KB - jmir_v24i5e27694_app6.pdf]

Multimedia Appendix 7
Supplementary File 7. Subgroup analysis for studies without deep learning.
[PDF File (Adobe PDF File), 424 KB - jmir_v24i5e27694_app7.pdf]

Multimedia Appendix 8
Supplementary File 8. Subgroup analysis for studies that used white light image.
[PDF File (Adobe PDF File), 506 KB - jmir_v24i5e27694_app8.pdf]

Multimedia Appendix 9
Supplementary File 9. Subgroup analysis for studies that used narrow band imaging techniques.
[PDF File (Adobe PDF File), 424 KB - jmir_v24i5e27694_app9.pdf]

Multimedia Appendix 10
Supplementary Table 1. Characteristics of the studies that compared diagnostic performance of artificial intelligence to endoscopists and its sensitivity analysis.
[PDF File (Adobe PDF File), 350 KB - jmir_v24i5e27694_app10.pdf]

Multimedia Appendix 11
Supplementary Table 2. Sensitivity analysis of the studies that included gastric lesions other than small gastric cancer lesions.
[PDF File (Adobe PDF File), 404 KB - jmir_v24i5e27694_app11.pdf]

Multimedia Appendix 12
Supplementary Table 3. Sensitivity analysis of the studies that do not detect early gastric cancer lesions based on pathological grading.
[PDF File (Adobe PDF File), 403 KB - jmir_v24i5e27694_app12.pdf]

Multimedia Appendix 13
Supplementary Table 4. Sensitivity analysis of the studies that separated training and testing data set during artificial intelligence training.
[PDF File (Adobe PDF File), 403 KB - jmir_v24i5e27694_app13.pdf]

Multimedia Appendix 14
Supplementary Table 5. Sensitivity analysis of the studies with low risk on index test.
[PDF File (Adobe PDF File), 577 KB - jmir_v24i5e27694_app14.pdf]

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018 Nov;68(6):394-424 [FREE Full text] [doi: 10.3322/caac.21492] [Medline: 30207593]
2. Ouyang G, Pan G, Liu Q, Wu Y, Liu Z, Lu W, et al. The global, regional, and national burden of pancreatitis in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. BMC Med 2020 Dec 10;18(1):388 [FREE Full text] [doi: 10.1186/s12916-020-01859-5] [Medline: 33298026]
3. Ford AC, Forman D, Hunt RH, Yuan Y, Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. BMJ 2014 May 20;348(may20 1):g3174-g3174 [FREE Full text] [doi: 10.1136/bmj.g3174] [Medline: 24846275]

XSL•FO
RenderX

4. Yaghoobi M, McNabb-Baltar J, Bijarchi R, Hunt RH. What is the quantitative risk of gastric cancer in the first-degree relatives of patients? A meta-analysis. World J Gastroenterol 2017 Apr 07;23(13):2435-2442 [FREE Full text] [doi: 10.3748/wjg.v23.i13.2435] [Medline: 28428723]
5. Rawla P, Barsouk A. Epidemiology of gastric cancer: global trends, risk factors and prevention. Prz Gastroenterol 2019;14(1):26-38 [FREE Full text] [doi: 10.5114/pg.2018.80001] [Medline: 30944675]
6. Ajani JA, Lee J, Sano T, Janjigian YY, Fan D, Song S. Gastric adenocarcinoma. Nat Rev Dis Primers 2017 Jun 01;3(1):17036. [doi: 10.1038/nrdp.2017.36] [Medline: 28569272]
7. Katai H, Ishikawa T, Akazawa K, Isobe Y, Miyashiro I, Oda I, Registration Committee of the Japanese Gastric Cancer Association. Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001-2007). Gastric Cancer 2018 Jan 17;21(1):144-154. [doi: 10.1007/s10120-017-0716-7] [Medline: 28417260]
8. Fock KM, Talley N, Moayyedi P, Hunt R, Azuma T, Sugano K, Asia-Pacific Gastric Cancer Consensus Conference. Asia-Pacific consensus guidelines on gastric cancer prevention. J Gastroenterol Hepatol 2008 Mar;23(3):351-365. [doi: 10.1111/j.1440-1746.2008.05314.x] [Medline: 18318820]
9. Choi KS, Jun JK, Park E, Park S, Jung KW, Han MA, et al. Performance of different gastric cancer screening methods in Korea: a population-based study. PLoS One 2012 Nov 29;7(11):e50041 [FREE Full text] [doi: 10.1371/journal.pone.0050041] [Medline: 23209638]
10. Yoshida N, Doyama H, Yano T, Horimatsu T, Uedo N, Yamamoto Y, et al. Early gastric cancer detection in high-risk patients: a multicentre randomised controlled trial on the effect of second-generation narrow band imaging. Gut 2021 Jan 02;70(1):67-75 [FREE Full text] [doi: 10.1136/gutjnl-2019-319631] [Medline: 32241898]
11. Pasechnikov V, Chukov S, Fedorov E, Kikuste I, Leja M. Gastric cancer: prevention, screening and early diagnosis. World J Gastroenterol 2014 Oct 14;20(38):13842-13862 [FREE Full text] [doi: 10.3748/wjg.v20.i38.13842] [Medline: 25320521]
12. Abbasi J. Artificial Intelligence Tools for Sepsis and Cancer. JAMA 2018 Dec 11;320(22):2303. [doi: 10.1001/jama.2018.19383] [Medline: 30535201]
13. Abbasi J. Artificial Intelligence-Based Skin Cancer Phone Apps Unreliable. JAMA 2020 Apr 14;323(14):1336. [doi: 10.1001/jama.2020.4543] [Medline: 32286633]
14. Abbasi J. Artificial Intelligence Improves Breast Cancer Screening in Study. JAMA 2020 Feb 11;323(6):499. [doi: 10.1001/jama.2020.0370] [Medline: 32044919]
15. Hwang TJ, Kesselheim AS, Vokinger KN. Lifecycle Regulation of Artificial Intelligence- and Machine Learning-Based Software Devices in Medicine. JAMA 2019 Dec 17;322(23):2285-2286. [doi: 10.1001/jama.2019.16842] [Medline: 31755907]
16. Matheny ME, Whicher D, Thadaney Israni S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. JAMA 2020 Feb 11;323(6):509-510. [doi: 10.1001/jama.2019.21579] [Medline: 31845963]
17. Rubin R. Artificial Intelligence for Cervical Precancer Screening. JAMA 2019 Feb 26;321(8):734. [doi: 10.1001/jama.2019.0888] [Medline: 30806677]
18. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
19. Voelker R. Cardiac Ultrasound Uses Artificial Intelligence to Produce Images. JAMA 2020 Mar 17;323(11):1034. [doi: 10.1001/jama.2020.2547] [Medline: 32181835]
20. Samoili S, Lopez Cobo M, Gomez Gutierrez E, De Prato G, Martinez-Plumed F, Delipetrev B. Defining Artificial Intelligence: Towards an operational definition and taxonomy of artificial intelligence. Luxembourg: Publications Office of the European Union; 2020.
21. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. Artificial Intelligence: A Modern Approach. Pearson Education, Inc; 2010:A.
22. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans Med Imaging 2016 May;35(5):1153-1159. [doi: 10.1109/tmi.2016.2553401]
23. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
24. Bang CS, Lee JJ, Baik GH. Artificial Intelligence for the Prediction of Helicobacter Pylori Infection in Endoscopic Images: Systematic Review and Meta-Analysis Of Diagnostic Test Accuracy. J Med Internet Res 2020 Sep 16;22(9):e21983 [FREE Full text] [doi: 10.2196/21983] [Medline: 32936088]
25. Ćirković A. Evaluation of Four Artificial Intelligence-Assisted Self-Diagnosis Apps on Three Diagnoses: Two-Year Follow-Up Study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
26. Liu T, Tsang W, Huang F, Lau OY, Chen Y, Sheng J, et al. Patients' Preferences for Artificial Intelligence Applications Versus Clinicians in Disease Diagnosis During the SARS-CoV-2 Pandemic in China: Discrete Choice Experiment. J Med Internet Res 2021 Feb 23;23(2):e22841 [FREE Full text] [doi: 10.2196/22841] [Medline: 33493130]

XSL•FO
RenderX

27. Shen J, Chen J, Zheng Z, Zheng J, Liu Z, Song J, et al. An Innovative Artificial Intelligence-Based App for the Diagnosis of Gestational Diabetes Mellitus (GDM-AI): Development Study. J Med Internet Res 2020 Sep 15;22(9):e21573 [FREE Full text] [doi: 10.2196/21573] [Medline: 32930674]
28. Gonçalves WG, Dos Santos MHP, Lobato FMF, Ribeiro-Dos-Santos Â, de Araújo GS. Deep learning in gastric tissue diseases: a systematic review. BMJ Open Gastroenterol 2020 Mar 26;7(1):e000371 [FREE Full text] [doi: 10.1136/bmjgast-2019-000371] [Medline: 32337060]
29. Luo H, Xu G, Li C, He L, Luo L, Wang Z, et al. Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study. The Lancet Oncology 2019 Dec;20(12):1645-1654. [doi: 10.1016/S1470-2045(19)30637-0] [Medline: 31591062]
30. Zhang X, Chen F, Yu T, An J, Huang Z, Liu J, et al. Real-time gastric polyp detection using convolutional neural networks. PLoS One 2019 Mar 25;14(3):e0214133 [FREE Full text] [doi: 10.1371/journal.pone.0214133] [Medline: 30908513]
31. Zhu Y, Wang Q, Xu M, Zhang Z, Cheng J, Zhong Y, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointest Endosc 2019 Apr;89(4):806-815.e1. [doi: 10.1016/j.gie.2018.11.011] [Medline: 30452913]
32. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995 Jan;117(1):167-178. [doi: 10.1037/0033-2909.117.1.167] [Medline: 7870860]
33. Song F, Khan KS, Dinnes J, Sutton AJ. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. Int J Epidemiol 2002 Feb;31(1):88-95. [doi: 10.1093/ije/31.1.88] [Medline: 11914301]
34. Namikawa K, Hirasawa T, Ikenoyama Y, Ishioka M, Tamashiro A, Shiroma S, et al. 343 Can Artificial Intelligence-Based Diagnostic System Perform Differential Diagnosis Of Gastric Cancer And Gastric Ulcer? In: Gastrointestinal Endoscopy. 2019 Jun Presented at: Digestive Disease Week 2019 American Society for Gastrointestinal Endoscopy Program; May 18-21, 2019; San Diego, CA p. AB74. [doi: 10.1016/j.gie.2019.04.047]
35. Cho B, Bang CS, Park SW, Yang YJ, Seo SI, Lim H, et al. Automated classification of gastric neoplasms in endoscopic images using a convolutional neural network. Endoscopy 2019 Dec;51(12):1121-1129. [doi: 10.1055/a-0981-6133] [Medline: 31443108]
36. Hori K, Takemoto S, Sakai Y. Automatic detection of early gastric cancer in endoscopic images using a transferring convolutional neural network. Conference Abstract. In: United European Gastroenterology Journal. 2018 Presented at: United European Gastroenterology Week 2018; October 21, 2018; Vienna, Austria p. A97. [doi: 10.1109/embc.2018.8513274]
37. Kanesaka T, Lee T, Uedo N, Lin K, Chen H, Lee J, et al. Computer-aided diagnosis for identifying and delineating early gastric cancers in magnifying narrow-band imaging. Gastrointest Endosc 2018 May;87(5):1339-1344. [doi: 10.1016/j.gie.2017.11.029] [Medline: 29225083]
38. Horiuchi Y, Aoyama K, Tokai Y, Hirasawa T, Yoshimizu S, Ishiyama A, et al. Convolutional Neural Network for Differentiating Gastric Cancer from Gastritis Using Magnified Endoscopy with Narrow Band Imaging. Dig Dis Sci 2020 May 04;65(5):1355-1363. [doi: 10.1007/s10620-019-05862-6] [Medline: 31584138]
39. Wu L, Zhou W, Wan X, Zhang J, Shen L, Hu S, et al. A deep neural network improves endoscopic detection of early gastric cancer without blind spots. Endoscopy 2019 Jun 12;51(6):522-531. [doi: 10.1055/a-0855-3532] [Medline: 30861533]
40. Ikenoyama Y, Hirasawa T, Ishioka M, Namikawa K, Yoshimizu S, Horiuchi Y, et al. Detecting early gastric cancer: Comparison between the diagnostic ability of convolutional neural networks and endoscopists. Dig Endosc 2021 Jan 02;33(1):141-150 [FREE Full text] [doi: 10.1111/den.13688] [Medline: 32282110]
41. Liu D, Gan T, Rao N, Xing Y, Zheng J, Li S, et al. Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process. Med Image Anal 2016 Aug;32:281-294. [doi: 10.1016/j.media.2016.04.007] [Medline: 27236223]
42. Yoon HJ, Kim S, Kim J, Keum J, Oh S, Jo J, et al. A Lesion-Based Convolutional Neural Network Improves Endoscopic Detection and Depth Prediction of Early Gastric Cancer. J Clin Med 2019 Aug 26;8(9):1310 [FREE Full text] [doi: 10.3390/jcm8091310] [Medline: 31454949]
43. Kubota K, Kuroda J, Yoshida M, Ohta K, Kitajima M. Medical image analysis: computer-aided diagnosis of gastric cancer invasion on endoscopic images. Surg Endosc 2012 May;26(5):1485-1489. [doi: 10.1007/s00464-011-2036-z] [Medline: 22083334]
44. Miyaki R, Yoshida S, Tanaka S, Kominami Y, Sanomura Y, Matsuo T, et al. Quantitative identification of mucosal gastric cancer under magnifying endoscopy with flexible spectral imaging color enhancement. J Gastroenterol Hepatol 2013 May 25;28(5):841-847. [doi: 10.1111/jgh.12149] [Medline: 23424994]
45. Yamakawa R, Harada M, Kawauchi K, Nyuzuki S, Masaya I. UEG Week 2018 Poster Presentations. 2018 Oct Presented at: United European Gastroenterology 2018; October 21, 2018; Vienna, Austria. [doi: 10.1177/2050640618792819]
46. Lui TKL, Guo C, Leung WK. Accuracy of artificial intelligence on histology prediction and detection of colorectal polyps: a systematic review and meta-analysis. Gastrointest Endosc 2020 Jul;92(1):11-22.e6. [doi: 10.1016/j.gie.2020.02.033] [Medline: 32119938]
47. Zhang Q, Wang F, Chen Z, Wang Z, Zhi F, Liu S, et al. Comparison of the diagnostic efficacy of white light endoscopy and magnifying endoscopy with narrow band imaging for early gastric cancer: a meta-analysis. Gastric Cancer 2016 Apr 29;19(2):543-552. [doi: 10.1007/s10120-015-0500-5] [Medline: 25920526]

XSL•FO
RenderX

48. Yu H, Wu L. 378 Randomized controlled trial of wisense, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. Gastrointestinal Endoscopy 2019 Jun;89(6):AB74-AB75. [doi: 10.1016/j.gie.2019.04.048]
49. Horiuchi Y, Hirasawa T, Ishizuka N, Tokai Y, Namikawa K, Yoshimizu S, et al. Performance of a computer-aided diagnosis system in diagnosing early gastric cancer using magnifying endoscopy videos with narrow-band imaging (with videos). Gastrointest Endosc 2020 Oct;92(4):856-865.e1. [doi: 10.1016/j.gie.2020.04.079] [Medline: 32422155]

Abbreviations

AI: artificial intelligence
AUC: area under the curve
HSROC: hierarchical summary receiver operating characteristic
NBI: narrow-band imaging
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies (revised)
SROC: summary receiver operating characteristic
WLI: white light imaging

Edited by R Kukafka; submitted 02.02.21; peer-reviewed by M Feng, S Pang; comments to author 11.05.21; revised version received 23.10.21; accepted 15.11.21; published 16.05.22.

Please cite as:
Chen PC, Lu YR, Kang YN, Chang CC
The Accuracy of Artificial Intelligence in the Endoscopic Diagnosis of Early Gastric Cancer: Pooled Analysis Study
J Med Internet Res 2022;24(5):e27694
URL: https://www.jmir.org/2022/5/e27694
doi:10.2196/27694
PMID:35576561

©Pei-Chin Chen, Yun-Ru Lu, Yi-No Kang, Chun-Chao Chang. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 16.05.2022. 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.

XSL•FO
RenderX

Original Paper

Impact of Hospital Characteristics and Governance Structure on the Adoption of Tracking Technologies for Clinical and Supply Chain Use: Longitudinal Study of US Hospitals

Xiao Zhu[1*], PhD; Youyou Tao[2*], PhD; Ruilin Zhu[3], PhD; Dezhi Wu[4], PhD; Wai-kit Ming[5], MPH, MD, PhD

[1]Department of Clinical Pharmacy and Pharmacy Administration, School of Pharmacy, Fudan University, Shanghai, China

[2]Department of Information Systems and Business Analytics, College of Business Administration, Loyola Marymount University, Los Angeles, CA, United States

[3]Management Science Department, Lancaster University Management School, Lancaster University, Lancaster, United Kingdom

[4]Department of Integrated Information Technology, College of Engineering and Computing, University of South Carolina, Columbia, SC, United States

[5]Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Sciences, City University of Hong Kong, Hong Kong, Hong Kong

[*]these authors contributed equally

Corresponding Author:
Ruilin Zhu, PhD
Management Science Department
Lancaster University Management School
Lancaster University
Bailrigg
Lancaster, LA1 4YX
United Kingdom
Phone: 44 1524592938
Email: ruilin.zhu@lancaster.ac.uk

Abstract

Background: Despite the increasing adoption rate of tracking technologies in hospitals in the United States, few empirical studies have examined the factors involved in such adoption within different use contexts (eg, clinical and supply chain use contexts). To date, no study has systematically examined how governance structures impact technology adoption in different use contexts in hospitals. Given that the hospital governance structure fundamentally governs health care workflows and operations, understanding its critical role provides a solid foundation from which to explore factors involved in the adoption of tracking technologies in hospitals.

Objective: This study aims to compare critical factors associated with the adoption of tracking technologies for clinical and supply chain uses and examine how governance structure types affect the adoption of tracking technologies in hospitals.

Methods: This study was conducted based on a comprehensive and longitudinal national census data set comprising 3623 unique hospitals across 50 states in the United States from 2012 to 2015. Using mixed effects population logistic regression models to account for the effects within and between hospitals, we captured and examined the effects of hospital characteristics, locations, and governance structure on adjustments to the innate development of tracking technology over time.

Results: From 2012 to 2015, we discovered that the proportion of hospitals in which tracking technologies were fully implemented for clinical use increased from 36.34% (782/2152) to 54.63% (1316/2409), and that for supply chain use increased from 28.58% (615/2152) to 41.3% (995/2409). We also discovered that adoption factors impact the clinical and supply chain use contexts differently. In the clinical use context, compared with hospitals located in urban areas, hospitals in rural areas (odds ratio [OR] 0.68, 95% CI 0.56-0.80) are less likely to fully adopt tracking technologies. In the context of supply chain use, the type of governance structure influences tracking technology adoption. Compared with hospitals not affiliated with a health system, implementation rates increased as hospitals affiliated with a more centralized health system—1.9-fold increase (OR 1.87, 95% CI 1.60-2.13) for decentralized or independent hospitals, 2.4-fold increase (OR 2.40, 95% CI 2.07-2.80) for moderately centralized health systems, and 3.1-fold increase for centralized health systems (OR 3.07, 95% CI 2.67-3.53).

Conclusions: As the first of such type of studies, we provided a longitudinal overview of how hospital characteristics and governance structure jointly affect adoption rates of tracking technology in both clinical and supply chain use contexts, which is

XSL•FO
RenderX

essential for developing intelligent infrastructure for smart hospital systems. This study informs researchers, health care providers, and policy makers that hospital characteristics, locations, and governance structures have different impacts on the adoption of tracking technologies for clinical and supply chain use and on health resource disparities among hospitals of different sizes, locations, and governance structures.

(J Med Internet Res 2022;24(5):e33742) doi:10.2196/33742

KEYWORDS

radio frequency identification; bar coding; tracking technology adoption; smart hospital; hospital affiliation; governance structure; location; clinical use; supply chain use

Introduction

Background

The extensive adoption of innovative tracking technologies has left almost no industry behind. Owing to strict health care laws, regulations, and policies, the health care industry has made great strides in the area, with a growing number of hospitals in the United States and worldwide beginning to reap the benefits of tracking technologies involved in, for example, optimizing health care processes, minimizing waste and human errors, and enhancing operational efficiency [1,2]. Upon approval, tracking technologies can be applied to enable health care providers to develop technology infrastructure in hospitals, resulting in greater efficiency in locating medications, patients, and staff in clinical processes and in tracking equipment, enhancing information sharing, and facilitating operations in the supply chain management process [1-5]. One of the main drivers in adopting tracking technologies is the meaningful use incentive program, which provides financial incentives for health care providers who use *certified* health technologies to meet a set of use criteria specified by the Centers for Medicare and Medicaid Services [6]. This program comprises 3 stages: stage 1 focuses on data capture and sharing, stage 2 relates to advanced clinical processes such as using tracking technologies for medication, and stage 3 concerns improved outcomes. The focus of our study is stage 2, where the adoption rate of tracking technologies has been increasing, given that the use of autotracking technologies to improve clinical processes has been one of the meaningful use core measures of stage 2, since 2012, effective in 2014 [7].

Of several applied instances in the field of tracking technology, barcodes and radio-frequency identification (RFID) are the most widely adopted tracking technologies [8]. Barcoding was introduced and used successfully in the health care industry 2 decades ago [7,9]. Linear or complex barcode technologies can encode patient, medicine, and asset information [9]. Unlike barcodes, which can only be read in *line of sight*, RFID has the advantage of using radio waves for automatic object identification, asset tracking, and data collection and transfer [10,11]. The implementation of RFID in the health care industry has been relatively recent and has become one of the major technological advancements in the future development of the health care sector [11,12]. For example, RFID can be applied with photosensors for smart hospitals to develop intelligent infrastructure, enabling better interactions between health care providers and patients and allowing more transparent and timely access to medical data [10]. As such, tracking technologies are essential for developing an intelligent infrastructure for current smart hospital systems, meaning that one of the goals of this study is to understand the factors involved in adopting tracking technologies (eg, barcoding and RFID).

Because of the appealing potential of tracking technologies to automate data, improve security, reduce counterfeiting and theft, and expedite and optimize clinical processes and supply chain management in the health care industry, their adoption in clinical and supply chain uses has been significantly outpaced by other widely adopted health technologies such as electronic health record systems [8,10-14]. Studies on the factors involved in health technology adoption are extensive in the medical, health informatics, and information systems literature, with hospital characteristics such as hospital size, teaching status, payer mix, accreditation status, ownership, and hospital affiliation being well documented as key factors influencing the implementation levels of various health technologies [15-19]. Tracking technologies such as RFID and barcodes have been carefully examined, as they possess the potential to automate and streamline processes in health care intelligence to improve clinical decision-making, patient care, and health care ecosystems for more accurate and efficient operations [20-22]. As such, tracking technologies, if adopted, can expedite and optimize clinical processes and health outcomes by tracking patients, health care workers, and hospital assets in real time [12,21,22], minimizing man-made mistakes and negligence [3,23-27] and delivering accurate information to health care systems, thereby unleashing the untapped benefits of digital innovations [3,22]. Together, they provide a solid foundation for efficient and effective health care practices, leading to more intelligent health care systems and operations [1,28]. Given the critical role and promising outlook of tracking technologies in health care, it is imperative to understand the key factors driving their adoption, which are already readily embedded in existing health care systems. However, to the best of our knowledge, no previous studies have systematically examined the critical factors that influence tracking adoption in clinical and supply chain use contexts, meaning that immediate investigation is required.

Within the umbrella of digital innovation, tracking technologies share some similarities, such as adopting digital features with other health information technologies (HITs). Nevertheless, they display a range of unique and distinctive characteristics that require thorough legal, clinical, and practical examination before adoption. First, unlike other HITs, concerns over privacy and security related to the use of tracking technologies are more prevalent and substantial [28], in that data associated with

XSL•FO
RenderX

tracking technologies are under tight restrictions and protection, as per the Health Insurance Portability and Accountability Act (eg, patient information including name, medical record number, and date of birth) [23]. In this regard, other factors such as time effects must be considered in the adoption process; understandably, little research has been undertaken to explore this because of the lack of available data. Second, previous studies largely considered HITs in the context of clinical use, incorporating technologies primarily used to aid clinical-focused processes, including capture, storage, and processing of clinical information, such as doctors' notes, patient records, and test results, as well as auxiliary systems for order entry and decision support [29]. In contrast, other HIT use contexts, such as HITs in supply chains, are generally offstage. Tracking technologies not only augment and optimize the capabilities of HITs in clinical use, speeding up clinical processes, for example, and facilitating supply chain management processes to be more cost-effective but also reduce unnecessary waste [28]. Therefore, the key factors that affect the adoption of HITs in the clinical process are likely to be distinctive from those contributing to the adoption of tracking technologies in supply chains, inviting further but separate investigations. For example, the type of hospital affiliation can impact supply chain use but not necessarily clinical use. Tracking technologies are also favorably envisioned as embedded in smart hospitals and Internet of Things–based hospitals as part of common intelligent health care initiatives to optimize health care processes, improve operational efficiency, and enhance safety for both patients and medical practitioners, particularly benefiting disadvantaged groups such as people who are disabled and older adults [11,30-33]. For instance, an assisted living system can locate and track people who are disabled and older adults, alert caregivers in real time in unexpected situations, and support daily activities for people who are disabled and older adults, including reminding them to take their daily medications safely [10,11,34]. Understanding the potential factors involved in the adoption of tracking technology pinpoints the complexities incorporated in taking full advantage of digital innovations, ranging from resource distribution to managerial, operational, and clinical practices central to achieving equitable and smart health care.

Objectives

One innovation in hospital management over the past few decades is strategic reconfiguration, which consolidates individual, unaffiliated hospitals into multihospital systems [35]. Previous studies have emphasized the importance of hospital systems in health technology adoption—hospitals affiliated with a health system can allocate more resources, improve coordination, increase market shares, and embrace greater financial performance and are more likely to adopt HITs, such as electronic medical record [19,35,36]. Hospitals affiliated with various health system types (eg, centralized vs independent health systems) may have different governance structures, indicating underlying mechanisms to ensure order in workflow management, including how work arrangements are structured (eg, structured within trustful networks, within various governance hierarchies, or as impersonal exchanges), and may have different service types (eg, centralized health systems with high levels of centralization of health services and decentralized health systems with high levels of decentralization of hospital services) [35]. This differentiation enables varying hospital work arrangements and health service types because of different governance structures and further affects technology adoption decisions. However, the general influence of governance structure on technology adoption in different use contexts in hospitals remains limited, and to the best of our knowledge, no study has examined this issue systematically based on a comprehensive data set. Our study, therefore, aims to extend the current understanding by identifying the relationship between hospital governance structure and tracking technology adoption.

Of the very limited number of quantitative studies previously undertaken to explore factors involved in the adoption of tracking technology in hospitals in the United States, Dey et al [28] conducted an empirical study in 2010 of 86 US hospitals, finding that both organizational and technological factors affect decisions to adopt RFID, whereas environmental factors such as uncertainty in a competitive market do not. Uy et al [7] examined the adoption trends of barcodes and RFID technologies with a larger data set of >5400 US hospitals from 2008 to 2012 and found that, in 2012, the adoption rate of barcodes for medication administration reached 58.1%, whereas adoption of RFIDs remained at 1.87%. Both witnessed their highest growth in adoption of medication administration in this period, compared with laboratory, pharmacy, and radiology use [7]. Of previous review studies on adoption of tracking technology, Wamba [4] conducted a comprehensive review of 22 articles published in the Journal of Medical Systems between 1997 and 2011 relating to application areas, types of benefits, and types of issues linked to RFID technology adoption. This review paper pointed out that the most highly published application area was that of patient management, the most widely discussed benefit was efficiency gain, and the most examined issues were data management, security, and privacy. These studies, however, were often limited by a small sample size [28] and considered only the early adoption period for tracking technologies (eg, 1997 to 2011 [4] and 2008 to 2012 [7]), reinforcing the notable gap in the longitudinal understanding of such tracking technologies adopted in health care for both clinical and supply chain use [1] and calling for more academic attention and further investigation.

This study, therefore, took the lead as the first longitudinal research study to empirically examine the different factors associated with the adoption of tracking technologies in different use contexts with more recent US hospital data sets. This was the first study to examine the impact of governance structure types on technology adoption in different use contexts in hospitals. Because of the complex nature of health care settings, we differentiate among the factors that influence the adoption of tracking technologies in the clinical and supply chain use contexts. Extant literature suggests that larger, urban, nonprofit, and teaching hospitals tend to possess more advanced resources, admit more complex patients with severe illnesses or multiple chronic conditions, and need to manage more complicated clinical workflows. When appropriate, these hospitals might implement a higher level of tracking technology to facilitate their clinical processes [37-39]. In response, we examined

XSL•FO
RenderX

whether hospital characteristics and locations would impact tracking technology adoption in a clinical context. Existing studies also posit that the more centralized the health system to which the hospital is affiliated, the more likely it is that the hospital has more centrally organized service delivery with higher incentives and resources such as revenue and personnel to enhance supply chain efficiency using tracking technologies [19,40]. As a result, we examined whether governance structure types would affect tracking technology adoption in the context of supply chain use. In summary, with a large US hospital-level longitudinal data set, we aim to (1) compare critical factors associated with the adoption of tracking technologies for clinical and supply chain uses and (2) examine how governance structure types affect the adoption of tracking technologies in different use contexts in hospitals.

Methods

Data and Sample

The data sets used in this study are obtained from 3 sources: the American Hospital Association's (AHA) annual surveys, the AHA's information technology (IT) supplemental files, and the US Bureau of Economic Analysis website. First, we collected data from the AHA's annual surveys to identify hospital characteristics and obtain health system data. Second, we used the AHA's IT supplemental files to capture the tracking technology implementation data. Third, we used data from the US Bureau of Economic Analysis website to obtain gross domestic product (GDP) per capita information [41]. The period from 2012 to 2015 was selected because, from 2012, the Centers for Medicare and Medicaid Services required hospitals to initiate tracking of medications using tracking technologies, including RFID or barcoding as part of meaningful use core measures [7]. Our final data set is a longitudinal one containing 3623 general medical and surgical hospitals from 50 states in the United States, from 2012 to 2015, of which 74.19% (2688/3623) of hospitals were observed at least twice.

Dependent Variables

The 2 dependent variables used in this study are tracking technology adoption for clinical use and supply chain use. We constructed tracking technology in a clinical use context by counting the number of technologies fully implemented and replacing paper record functionality at a hospital, an approach widely used in information systems and health care literature [29,42]. Implementation levels of tracking technology for each itemized technology are measured on a 6-point scale, where 1 indicates *fully implemented across all units*, 2 to 5 indicates *partially implemented* at different levels, and 6 indicates *not in place and not considering implementing.* To calculate the implementation level for each technology function, we applied a coding scheme to separate full implementation—technology has completely replaced paper record functionality—from partial or no implementation. We retained the original coding of 1 as 1 and then coded the responses between 2 and 6 as 0. There are 4 tracking technology functions in the clinical use context: medication administration, patient verification, caregiver verification, and pharmacy verification. Therefore, the resultant tracking technology implementation level in a clinical use context ranged from 0 to 4. On this scale, 0, 1, 2, 3, and 4 indicate that the hospital fully implemented none, one, two, three, and four of the four tracking technologies, respectively. We then applied a binary variable to code completely implemented tracking technology in a clinical context across all 4 technology functions, for which 1 indicates fully implemented across all *four* technology functions and 0, otherwise. We also used a binary variable to code the tracking technology for supply chain management, as there is only 1 technology unit in this variable, for which 1 indicates fully implemented and 0, otherwise. For example, if a hospital has completely digitalized its tracking process using tracking technologies for clinical use, including medication administration, patient verification, caregiver verification, pharmacy verification, and supply chain management, the hospital's clinical and supply use adoption will be coded as 1.

Independent Variables

We included 3 sets of independent explanatory variables. The first set of variables was related to hospital characteristics, such as hospital size, ownership, and teaching status. *Hospital size* was measured based on the number of staffed beds. *Hospital ownership status* was coded as a binary variable denoting whether the hospital was a for-profit hospital. *Teaching status* was also coded as a binary variable, where 1 indicated a teaching hospital and 0, otherwise. We defined teaching hospitals as members of the Council of Teaching Hospitals of the Association of American Medical Colleges. The second set included the hospital location variables. *Hospital location* was measured using 3 dummy variables: metropolitan, micropolitan, and rural regions. We also measured the *state economic condition* where the hospital is located because previous studies found that per capita GDP plays an important role in technology adoption and use [43,44]. Thus, we assume that such a condition would be linked to the adoption of digital innovations in health care, which is worth examining in the context of tracking technology. Per capita GDP was measured using GDP per capita per state. We first ranked hospitals from high to low based on their state GDP per capita. We then coded hospitals belonging to the first half as *economic leading state* and those in the second half as *economic leveling state.* The third set comprises hospital governance structures because previous research indicates that governance structure is significantly associated with technology adoption [19,40]. *Governance structure* is measured according to whether the hospital is affiliated with a health system and, if so, the level of centralization in multihospital systems. Centralized health systems have unified asset ownership of affiliated hospitals and offer an array of products and services [45]. As in previous research [19,40], we applied 5 dummy variables to measure governance structure based on the diversification of health services and products and centralization of authority across health systems (out of health system as the base category) [45]. These variables define whether hospitals are (1) in centralized health systems with high levels of centralization of hospital service delivery, physician arrangements, and insurance product development; (2) in health systems with highly centralized physician arrangements and insurance product development; (3) in moderately centralized health systems with both centralized and decentralized activities

XSL•FO
RenderX

for hospital services, physician arrangements, and insurance product development; (4) in decentralized health systems with a high degree of decentralization of hospital services, physician arrangements, and insurance product development; and (5) in independent hospital systems with limited differentiation among hospital services, physician arrangements, and insurance product development.

Statistical Analysis

To examine the factors involved in the adoption of tracking technology in both clinical and supply chain use contexts in US hospitals, we used a mixed effects model using a population approach. This model is an extension of the simple fixed effects modeling to account for both fixed and random effects. This is particularly useful when data violate the independence assumption that arises from a hierarchical structure. For example, in this study, there were 2 levels: between hospitals (level 1) and within hospitals (level 2). As the data records for this study were collected from 3623 hospitals over 4 years, the source of variability in the observations can be attributed to either within-hospital or between-hospital effects. Repeated observations over the years from the same hospital are subject to hospital-level time-invariant unobserved effects, as within a given hospital, records are more similar. The units sampled at the highest level (ie, hospitals in this study) were independent. As our 2 dependent variables—tracking technology adoption for clinical use and tracking technology adoption for supply chain use—are binary variables, we developed a mixed effects population logistic regression model to examine the relationships among the adoption of tracking technologies (ie, clinical use vs supply chain use), hospital characteristics, and governance structure with the adjustment of time effect. Nonlinear mixed effects modeling software (NONMEM, version 7.5.0; ICON Development Solutions) was used for the modeling [46]. The Laplace estimation method was applied for parameter estimation. Perl Speaks NONMEM (PsN 4.8.0; Department of Pharmacy, Uppsala University) was used for model diagnostics and R (version 3.5.3) was used for data exploration before modeling and postprocessing of the results [29].

Initially, correlations among the covariates were explored. Exploratory graphical and statistical evaluations were performed to identify the relationship between estimated individual random effects and covariates. ANOVA tests for categorical covariates and linear regression for continuous covariates were used to identify possible univariate covariate relationships at $P<.05$. Only after statistically significant covariates were identified was the association between relevant hospital covariates and tracking technology adoption parameters further evaluated using a forward inclusion and backward elimination strategy, with model selection carried out using a log likelihood ratio test at an acceptance P value of .01 (a decrease in objective function value>6.63) in the forward step and a P value of .001 (an increase in objective function value >10.83) in the backward step. The final selection of relevant covariates was based on statistical evidence and health technology knowledge and interpretation. The derived model was further refined based on model convergence, parameter precision and impact of the covariate effect. The predictive performance of the final population logistic regression model was evaluated using visual predictive check (VPC) plots. Plots of observed data distributions were compared with simulated distributions to demonstrate the model's ability to adequately predict data on which the model is based. VPCs were based on 1000 simulations and stratified by the covariates of potential interest.

Results

Overview

A total of 3623 hospitals in 50 states in the United States, from 2012 to 2015, were included in this study (the complete list of hospitals can be accessed in Multimedia Appendix 1). Of these 3623 hospitals, 928 (25.61%) were in rural areas, 3133 (86.48%) were nonprofit hospitals, 223 (6.16%) were teaching hospitals, and 2158 (59.56%) were affiliated with health systems, and the mean total of the number of beds was 174 (SD 201). Detailed demographics of the included hospitals are listed in Table 1, and Table 2 presents the results of the adoption of tracking technologies over time. The AHA IT supplement survey specifies fully implemented as the status of technology that has completely replaced paper record functionality. In this regard, from 2012 to 2015, as per the data set, the proportion of hospitals that have fully adopted tracking technologies in digitalized medication administration, patient verification, caregiver verification, and pharmacy verification in a clinical use context increased from 36.34% (782/2152) to 54.63% (1316/2409), whereas the proportion of hospitals that have fully adopted tracking technologies to digitalize supply chain management increased from 28.58% (615/2152) to 41.3% (995/2409), demonstrating that the tracking functionality for both clinical use and supply chain use has been increasingly digitalized in this period.

XSL•FO
RenderX

Table 1. Demographic information from the included hospitals (N=3623).

Demographics	Overall
Location, n (%)	
Metro	2019 (55.72)
Micro	676 (18.65)
Rural	928 (25.61)
Profit, n (%)	
Not-for-profit	3133 (86.47)
For-profit	490 (13.52)
Teaching hospital, n (%)	
Yes	223 (6.15)
No	3400 (93.84)
State economic condition[a], n (%)	
Economic leveling state	1753 (48.38)
Economic leading state	1870 (51.61)
Governance structure: type of hospital affiliation (HS)[b], n (%)	
Centralized HS	310 (8.55)
Centralized physician and insurance HS	54 (1.49)
Moderately centralized HS	276 (7.61)
Decentralized HS	1419 (39.16)
Independent HS	99 (2.73)
Within HS	2158 (59.56)
Out of HS	1465 (40.43)
Total bed count, mean (SD)	174 (201)

[a]Economic leading state: top 25 states in gross domestic product per capita; economic leveling state: last 25 states in gross domestic product per capita.
[b]HS: health system.

Table 2. Adoption of tracking technologies in the United States from 2012 to 2015.

Usage	Tracking technologies year, n (%)			
	2012 (N=2152)	2013 (N=2012)	2014 (N=2277)	2015 (N=2409)
Clinical use				
Fully implemented	782 (36.33)	892 (44.33)	1190 (52.26)	1316 (54.62)
Not fully implemented	1370 (63.66)	1120 (55.66)	1087 (47.73)	1093 (45.37)
Supply				
Fully implemented	615 (28.57)	746 (37.07)	909 (39.92)	995 (41.3)
Not fully implemented	1537 (71.42)	1266 (62.92)	1368 (60.07)	1414 (58.69)

Tracking Technologies for Clinical Use

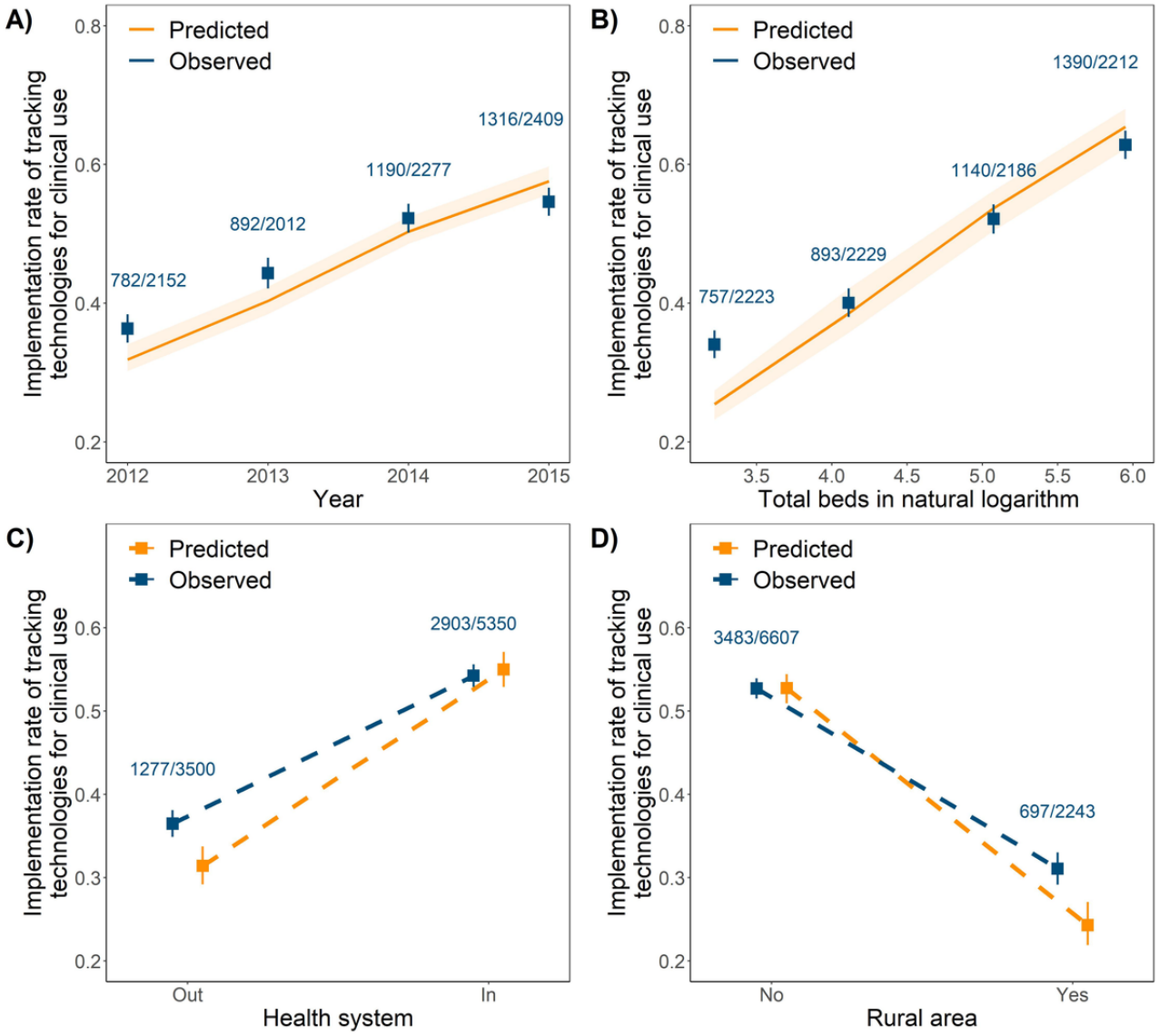
As shown in the VPC plots (Figure 1), the mixed effects population logistic regression model developed could well describe the adoption of tracking technologies for clinical use. All the population parameters for a typical hospital (defined as a hospital with 101 beds, not affiliated to a health system, and not in a rural area) were precisely estimated: the intercept was estimated to be −1.08 (relative SE 8%), and the slope was estimated to be 0.369 (relative SE 8%; Table 3). The total beds in natural logarithm, rural locations, and health systems were statistically significant covariates on the intercept. The relative univariate effects of total beds, rural locations, and health systems on the implementation rate of tracking technologies for clinical use are summarized as a forest plot in Figure 2.

The model developed has the potential to predict the increasing trend in the implementation rate of tracking technologies in clinical use over a period of years (Figure 1A). A positive relationship was identified with hospital size (reflected by the

XSL•FO
RenderX

total number of beds; Figure 1B). Similarly, the implementation rate increased by a median of 1.7-fold for hospitals affiliated with the health system relative to those that were not affiliated (Figure 1C and Figure 2). These results imply the influence of hospital infrastructure (both physical and organizational structures) on the adoption of tracking technologies in clinical use. Meanwhile, the implementation rate decreased by a median of 32% in hospitals located in rural areas relative to those in urban areas, showing clear evidence of location disparity (Figure 1D and Figure 2).

Figure 1. Visual predictive check plots of final population logistic regression model for the adoption of tracking technologies for clinical use over time. (A) the influence of time on the implementation rate of tracking technologies for clinical use; (B) the influence of total beds on the implementation rate of tracking technologies for clinical use; (C) the influence of health system on the implementation rate of tracking technologies for clinical use; (D) the influence of location (in the rural area or not) on the implementation rate of tracking technologies for clinical use. The blue dots show observed implementation rate; the blue error bars indicate a 95% CI in the observed implementation rate; the yellow dots and yellow solid lines show the median implementation rate from model prediction; the yellow error bars and the yellow area indicate a 95% prediction interval for the implementation rate.



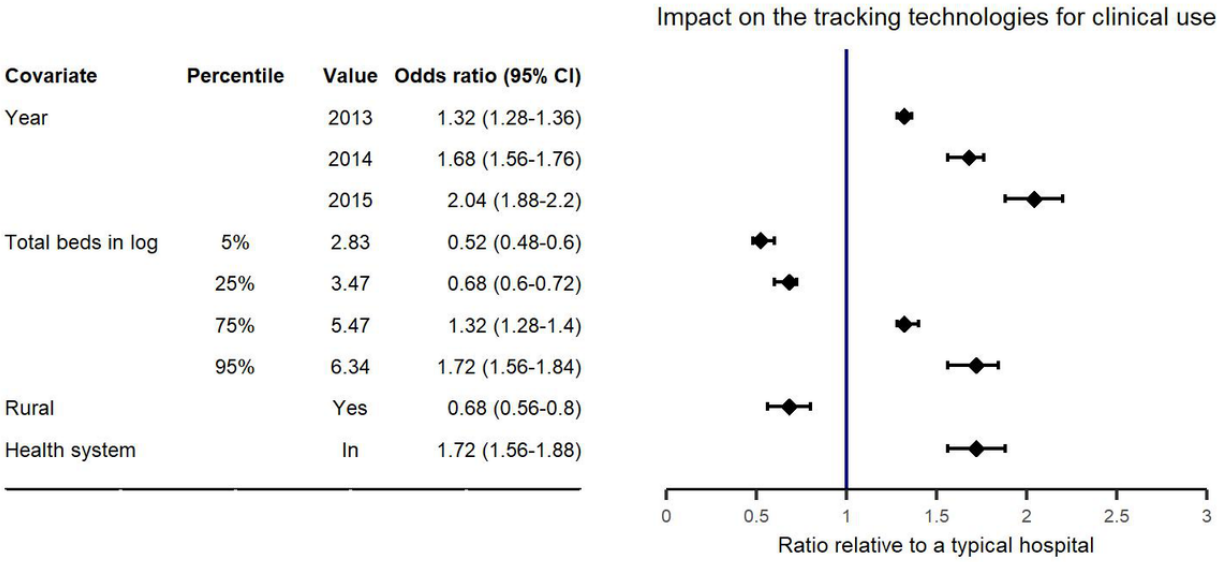
XSL•FO
RenderX

Table 3. Parameter estimates of final population logistic regression model for the adoption of tracking technologies for clinical use.

	Estimate (relative SE; %)
Fixed effects	
Intercept	−1.08 (8)
Time effect	0.369 (8)
Log total bed	0.452 (10)
Rural area	−0.535 (21)
Health system	0.79 (11)
Random effects	
Intercept	2.55 (8)
Time effect	0.11 (47)



Figure 2. Forrest plot of covariate effects on the implementation rate of tracking technologies for clinical use. The solid vertical line corresponds to a ratio of 1 and represents a typical hospital. Points and whiskers represent the estimate and 95% CI, respectively. A typical hospital is defined as a hospital with a total of 101 beds, not part of a health system, and not in a rural area in 2012.



Tracking Technologies for Supply Chain Use

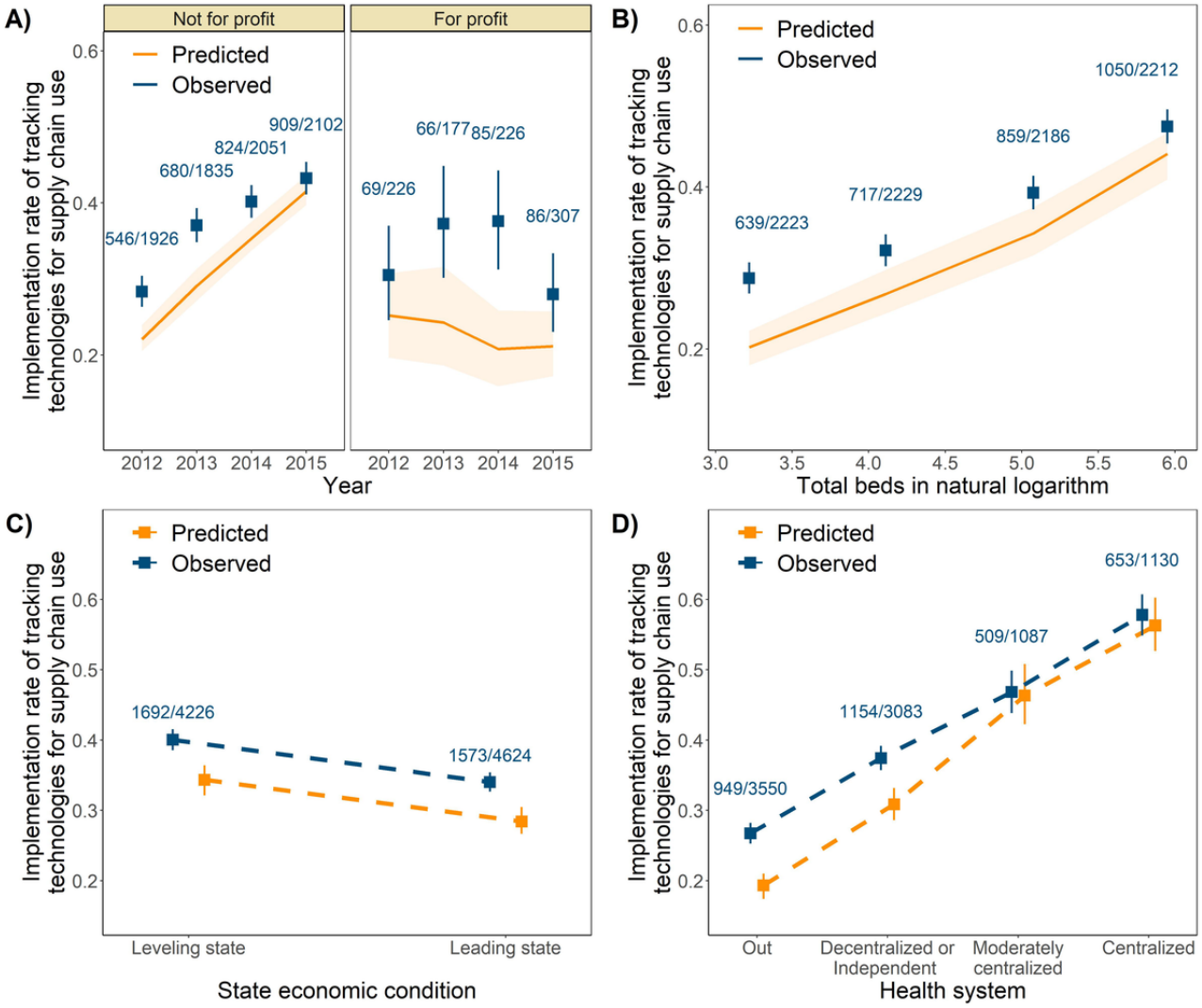
As shown in the VPC plots (Figure 3), the mixed effects population logistic regression model developed could well describe the adoption of tracking technologies for supply chain use. All population parameters for a typical hospital (defined as a nonprofit hospital with 101 beds and not affiliated with a health system in an economic leveling state) were precisely estimated: the intercept was estimated to be −1.72 (relative SE 6%), and the slope was estimated to be 0.3 (relative SE 10%; Table 4). Total beds in natural logarithm, state economic condition, and affiliation to a health system were found to be statistically significant covariates on the intercept, and profit status was found to be a statistically significant covariate on the slope. The relative univariate effects of total beds, state economic condition, and type of hospital affiliation on the implementation rate of tracking technologies in supply chain use are summarized as a forest plot in Figure 4.

The model developed can also predict the increasing trend in the implementation rate of tracking technologies for supply chain use over a period of 4 years in not-for-profit hospitals, as well as stagnation in development among hospitals running for profit (Figure 3A). This indicates that for-profit hospitals are more reluctant to implement these new technologies. The implementation rate of tracking technologies for supply chain use grew in parallel with increasing hospital size (as reflected by the total number of beds in the hospital; Figure 3B).

XSL•FO
RenderX

Furthermore, implementation rates increased as hospitals affiliated with a more centralized health system—1.9-fold increase (odds ratio [OR] 1.87, 95% CI 1.60-2.13) for decentralized or independent hospitals, 2.4-fold increase (OR 2.40, 95% CI 2.07-2.80) for moderately centralized health systems, and 3.1-fold increase for centralized health systems (OR 3.07, 95% CI 2.67-3.53), compared with hospitals not affiliated with a health system (Figure 3D and Figure 4). When compared with tracking technologies for clinical use, these results demonstrate a similar impact of hospital infrastructure on the adoption of tracking technologies for supply chain use: hospitals with better infrastructure tend to be more responsive in adopting tracking technologies. Surprisingly, the implementation rate decreased by a median of 30% in hospitals in economic leading states relative to those in economic leveling states (Figures 3C and Figure 4).

Figure 3. Visual predictive check plots of final population logistic regression model for the adoption of tracking technologies for supply chain use over time. (A) the influence of time on the implementation rate of tracking technologies for supply chain use; (B) the influence of total beds on the implementation rate of tracking technologies for supply chain use; (C) the influence of state economic condition on the implementation rate of tracking technologies for supply chain use; (D) the influence of health system on the implementation rate of tracking technologies for supply chain use. The blue dots show observed implementation rate; the blue error bars indicate a 95% CI in the observed implementation rate; the yellow dots and yellow solid lines show the median implementation rate from model prediction; the yellow error bars and the yellow area indicate a 95% prediction interval in the implementation rate.



XSL•FO
RenderX

Table 4. Parameter estimates of final population logistic regression model for the adoption of tracking technologies for supply chain use.

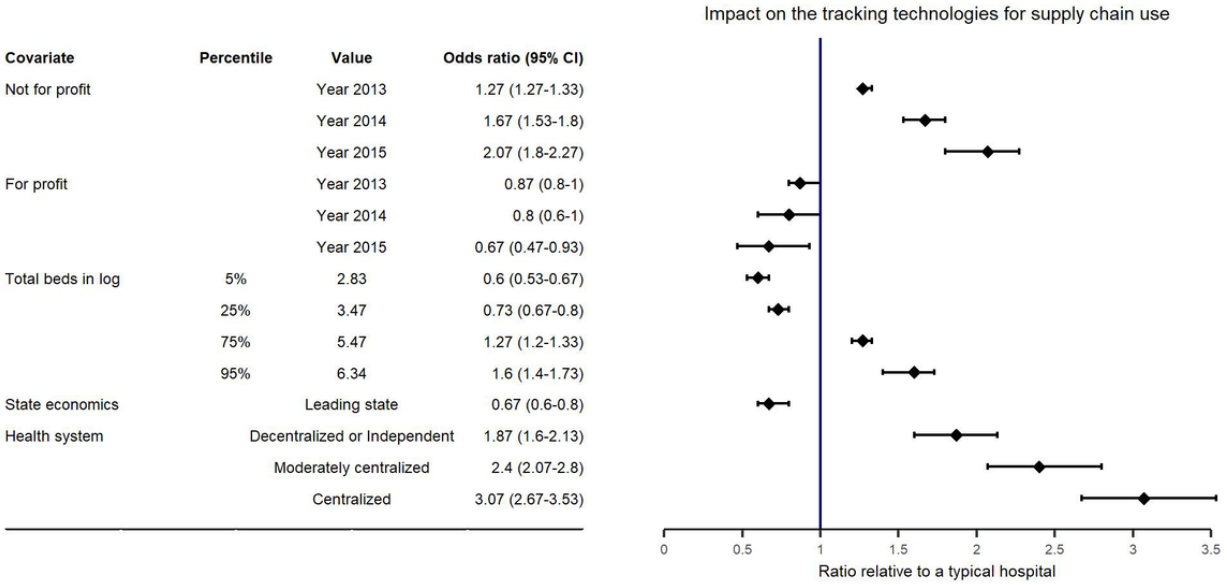
	Estimate (relative SE; %)
Fixed effects	
Intercept	−1.72 (6)
Time effect	0.3 (10)
Log total beds	0.321 (12)
Economic leading state	−0.428 (20)
Centralized HS[a]	1.57 (9)
Moderately centralized HS	1.16 (11)
Decentralized or independent HS	0.772 (13)
Run for-profit effect on time effect	−1.48 (15)
Random effects	
Intercept	3.22 (8)
Time effect	—[b]

[a]HS: health system.

[b]Data does not support the inclusion of random effect on time effect.



Figure 4. Forrest plot of covariate effects on implementation rate of tracking technologies for supply chain use. The solid vertical line corresponds to a ratio of 1 and represents a typical hospital. Points and whiskers represent the estimate and 95% CI, respectively. A typical hospital is defined as a not-for-profit hospital with a total of 101 beds, not part of a health system, and in an economic leveling state in 2012.



Discussion

Principal Findings

With a large US hospital-level longitudinal data set, we observed that, from 2012 to 2015, the proportion of hospitals in which tracking technologies were fully implemented for clinical use increased from 36.34% (782/2152) to 54.63% (1316/2409) and for supply chain use increased from 28.58% (615/2152) to 41.3% (995/2409). We found that larger hospitals were more likely to fully adopt tracking technologies in both clinical and supply chain use contexts, indicating health resource disparities among hospitals of different sizes. We also discovered that adoption factors affect the clinical and supply chain use contexts differently. In the clinical use context, compared with hospitals located in urban areas, hospitals in rural areas (OR 0.68, 95% CI 0.56-0.80) are less likely to fully adopt tracking technologies,

XSL•FO
RenderX

showing evidence of location disparity. In the context of supply chain use, the type of governance structure influences tracking technology adoption. Compared with hospitals not affiliated with a health system, implementation rates increased as hospitals affiliated with a more centralized health system—1.9-fold increase (OR 1.87, 95% CI 1.60-2.13) for decentralized or independent hospitals, 2.4-fold increase (OR 2.40, 95% CI 2.07-2.80) for moderately centralized health systems, and 3.1-fold increase for centralized health systems (OR 3.07, 95% CI 2.67-3.53).

Comparison With Previous Work

Given that studies on the adoption of tracking technologies have lagged in general health technology adoption studies, and studies undertaken are either limited by a small sample size or subject to early adoption periods, we attempted to fill this gap by applying a census data set from 2012 to 2015 to examine the factors involved in tracking technology adoption in both clinical and supply chain use contexts. Using mixed effects population logistic regression models, we identified several hospital characteristics and governance structure factors associated with tracking technology adoption. Consistent with previous studies on the impact of hospital size on technology adoption [15,47], our results show that larger hospitals are more likely to adopt tracking technologies in both the clinical and supply chain use contexts. In addition to considering hospital size, we found that hospitals in health systems are more likely to adopt tracking technologies in both clinical and supply chain use contexts. One reason for this is that tracking technology implementation cannot be accomplished in a single stroke. It requires the integration of tags, readers, networks, databases, systems, and expertise from different domains including RFID and barcode technology, medicine, information systems development, telecommunications, and systems integration [44]. Tracking technology is also part of the hospital technology infrastructure because it allows hospitals to capture, store, and streamline data and processes and can be integrated with other HITs such as electronic health records [48]. As infrastructure technology, the benefit of tracking technology adoption may only be realized in the long term. Thus, large hospitals or those within health systems urgently need to manage more complex patients with multiple chronic conditions with sufficient resources and capabilities to implement tracking technology and process large volumes of real-time data generated by the tracking technology.

In the context of clinical use, our results supplement existing studies with additional findings, identifying that rural hospitals are less likely to adopt tracking technologies. One possible reason is that, in contrast to hospitals located in metropolitan and micropolitan areas, those in rural areas tend to accept patients with less severe and less complicated diseases, which are more easily diagnosed and treated by local health care providers, thus requiring less sophisticated technology for clinical use [38]. Another possible reason is that rural hospitals may have fewer available resources to promote and implement new HIT functions. For example, urban and suburban hospitals are found to have wide access to experienced technical support staff to assist with the implementation process for new HIT functions, resulting in more gains from the adoption of new HIT functions [49], whereas such high-caliber personnel may not be readily available for rural hospitals.

In the context of supply chain use, our results show that compared with not-for-profit hospitals, for-profit hospitals are less likely to adopt tracking technologies. Our results, shown in Figure 3A, indicate that the implementation rate of tracking technologies for for-profit hospitals increases and then decreases (inverted U-shaped implementation rate) over time but increases over time for not-for-profit hospitals. One possible reason is that for-profit hospitals tend to pursue a high return on investments and thus often invest more in profitable services and avoid less profitable investments [50]. Given that supply chain management is indirectly related to hospital revenue, the use of tracking technologies in supply chains may not be prioritized in their investment lists against budgets. Thus, only for-profit hospitals that are highly motivated to reap the benefits of tracking technology (eg, to enhance efficiency) and obtain financial incentives from meaningful use fully implemented tracking technologies immediately after the addition of the autotracking medication requirement as a core measure in meaningful use in 2012, explaining why the implementation rate increased sharply from 2012 to 2013 for for-profit hospitals and the acceleration rate of implementation decreased from 2013 to 2015 [6]. A surprising finding also emerges from research that hospitals in economic leading states are less likely to adopt tracking technologies for supply chain use. One plausible reason is that economic leading states are generally more populous and thus, these urban hospitals need to attend to larger numbers of patients, thereby putting their funding priorities more on efficiency enhancement for immediate patient treatment, involving featured clinical processes rather than managerial operations, such as supply chain management.

Our study also extends the current understanding of how governance structure influences technology adoption by identifying the relationship between governance structure types and tracking technology adoption. We revealed that hospitals affiliated with health systems are more likely to adopt tracking technologies for clinical use, whereas types of hospital affiliation do not affect the adoption of tracking technologies for clinical use. We also find that the type of hospital affiliation affects the adoption of tracking technologies for supply chain use—hospitals affiliated with more centralized health systems are more likely to adopt tracking technologies for supply chain use. Compared with other types of hospital affiliations (eg, decentralized or independent or moderately centralized), centralized systems provide a higher percentage of their services at the system level, making them more likely to have higher incentives to increase supply chain efficiency using tracking technologies and develop the long-term tracking technology–related infrastructure of smart hospitals [40]. In addition, tracking technologies adopted for supply chain use, compared with tracking technologies for clinical use, may be costlier, complicated, and take longer to implement, requiring systematic and strategic planning, implementation, and integration and a more centralized health system.

Overall, 3 implications are set out in our study for researchers, health care stakeholders, and policy makers. First, our study indicates that the context of technology use (ie, clinical use or

XSL•FO
RenderX

supply chain use) influences the tracking of technology adoption. For example, we found that for supply chain use, governance structure types are important factors in the adoption of tracking technologies, but this is not the case for clinical use. In other words, there is no one-size-fits-all solution for adopting tracking technologies in the field of health care. When examining the impact of tracking technologies, practitioners, both academic and practical, should develop a holistic view of the adoption context and cannot assume that related factors can be generalized from other contexts. Health care practitioners who aspire to establish tracking technology–enabled (eg, RFID-enabled) smart hospitals, for example, are in favor of implementing tracking technologies for clinical use, facilitating information sharing, patient identification, and medical equipment tracking, and in supply chains to avoid drug counterfeiting and to enhance supply chain operations [11]. However, our study shows that the factors involved in the adoption of tracking technologies for clinical and supply chain use may be different. Understanding differences in adopting tracking technologies in various use contexts will help all hospitals involved in health care to plan and implement tracking technologies more strategically and avoid any possible pitfalls while maximizing their benefits from the outset. When tracking technologies are further leveraged in conjunction with other technologies, such as electronic health records, electronic data interchange technologies, mobile devices, and telehealth, caution over the context in question may still be relevant, suggesting that it is important for future studies to examine the different use contexts (eg, clinical and supply chain use) of tracking technologies, as highlighted in and demonstrated by our study.

Second, similar to initial studies that examine the effects of governance structure on longitudinal tracking technology adoption, our results suggest that the impact of governance structure types should be emphasized in technology adoption studies and that the underlying mechanisms require further investigation. For example, we identified that hospitals affiliated with more centralized health systems are more likely to adopt tracking technologies for supply chain use because of the centralized hospital structure settings, allowing resources to be prioritized and allocated to improve operational efficiency for more efficient and streamlined use, thus serving larger patient populations with personalized medicine. This feasibly occurs when systematic integration and synchronization for various solo practices are implemented in centralized smart hospital systems. Future studies are required to investigate the underlying mechanisms (ie, managerial support) linking technology adoption and governance structure and examine whether the findings of this study can be extended to other technology innovations.

Third, our results suggest that disparities may exist in health resources between hospitals of various sizes and governance structures. We found that larger hospitals and hospitals affiliated with health systems, especially more centralized health systems, are more likely to adopt tracking technologies. Compared with small and independent hospitals, these hospitals tend to have more human and financial resources to become the first adopters of advanced technologies. A potentially uneven distribution should be given ample attention before the trend becomes so established that it compounds the already sizable digital gap among different types of hospitals [51]. This is extremely important for smart hospital development, as tracking technologies can be applied as a technology infrastructure in smart ecosystem design. Given the increasingly important role of tracking technologies in transforming existing health care providers into smart hospitals, understanding the key factors involved in tracking technology adoption provides governments with evidence-based findings, supporting them to develop more feasible quantified health resources with an allocation scheme that promotes barcodes, RFID-enabled smart hospitals, and equitable health care. Thus, our study highlights the need for up-to-date government policies related to reasonable resource allocation for tracking technology implementation and its use in establishing and developing smart hospitals, including the use of tracking technologies in patient care, drug management, security and privacy, and tailored interventions from regulatory bodies or policy makers.

Strengths

This study is the first longitudinal research to empirically examine the different factors associated with the adoption of tracking technologies in different use contexts. This is also the first study to examine the impact of governance structure types on technology adoption in different use contexts in hospitals. In doing so, we provided a census assessment and longitudinal overview of how hospital characteristics and governance structure are related to the adoption rates of tracking technology in both clinical and supply chain use contexts. This study informs researchers, health care providers, and policy makers that hospital characteristics, locations, and governance structures have different impacts on the adoption of tracking technologies for clinical and supply chain use and on health resource disparities among hospitals of different sizes and with different locations and governance structures. This study has important managerial implications for the development of smart hospitals using tracking technologies to establish their hospital infrastructure and practical implications for examining the impact of governance structure types on the adoption of other technologies in health contexts.

Limitations and Future Directions

This study had some limitations. First, as comprehensive as the data set was, the timeframe was limited to the period from 2012 to 2015. Despite our rationale to address the scarcity of research into health care tracking technology by combing through details related to the issue of tracking technology adoption since its initial implementation in 2012 for the second stage of meaningful use, we caution that further development could have been in place as part of recent uptakes. Thus, it is necessary to conduct this research in conjunction with additional data. Second, we put in place 2 application scenarios to examine tracking technology in the clinical and supply chain use contexts. However, this examination has the potential for a more detailed focus on capturing additional particulars. For example, future research could examine the factors that influence the implementation of different clinical uses of tracking technologies, such as medication administration, patient verification, caregiver verification, and pharmacy verification.

XSL•FO
RenderX

Conclusions

This study provides a census assessment of the adoption of both clinical and supply chain tracking technologies in US hospitals and offers a comprehensive overview of the hospital characteristics and governance structure associated with tracking technology adoption. From an academic perspective, this study unearths the staggered adoption of health tracking technology in hospitals in various categories, suggesting that hospital characteristics and governance structures have a significant impact on the implementation level and rate of tracking technology in clinical and supply chain use. It expands our understanding of digital innovations in health care, providing further evidence relating to tracking technology and outlining implications that can be leveraged from a managerial point of view. This study informs health care providers and policy makers of the possible guidance references that tailored policies should be in place to further promote the ongoing digital transformation in health care, as hospital characteristics and governance structures have different influences on the digitalization process. These outcomes can facilitate both academics and practitioners in putting forward future research to further reveal the nature and scope of tracking technology in developing smart hospitals and personalized health care in general.

Acknowledgments

The authors thank Dr Jing Yuan from the School of Pharmacy at Fudan University for her suggestions on this manuscript. This work was funded by the Scientific Research Foundation for Talented Scholars at Fudan University (JIF301052).

Data Availability

The data sets analyzed during this study are not publicly available owing to the restrictions of use from the American Hospital Association data use agreement but can be accessed with permission from the American Hospital Association [52].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Complete list of hospitals applied in this study.
[XLSX File (Microsoft Excel File), 72 KB - jmir_v24i5e33742_app1.xlsx]

References

1. Gu VC, Black K. Integration of TTF and network externalities for RFID adoption in healthcare industry. Int J Product Perform Manag 2020 Mar 17;70(1):109-129. [doi: 10.1108/ijppm-11-2018-0418]
2. Rosenbaum BP. Radio frequency identification (RFID) in health care: privacy and security concerns limiting adoption. J Med Syst 2014 Mar;38(3):19. [doi: 10.1007/s10916-014-0019-z] [Medline: 24578170]
3. Coustasse A, Tomblin S, Slack C. Impact of radio-frequency identification (RFID) technologies on the hospital supply chain: a literature review. Perspect Health Inf Manag 2013 Oct 1;10(Fall):1d [FREE Full text] [Medline: 24159272]
4. Fosso Wamba S. RFID-enabled healthcare applications, issues and benefits: an archival analysis (1997-2011). J Med Syst 2012 Dec;36(6):3393-3398. [doi: 10.1007/s10916-011-9807-x] [Medline: 22109670]
5. Reyes PM, Li S, Visich JK. Determinants of RFID adoption stage and perceived benefits. Eur J Oper Res 2016 Nov;254(3):801-812. [doi: 10.1016/j.ejor.2016.03.051]
6. Stage 2 overview tipsheet. Centers for Medicare & Medicaid Services. 2012. URL: https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/downloads/stage2overview_tipsheet.pdf [accessed 2021-07-23]
7. Uy RC, Kury FP, Fontelo PA. The state and trends of barcode, RFID, biometric and pharmacy automation technologies in US hospitals. AMIA Annu Symp Proc 2015 Nov 5;2015:1242-1251 [FREE Full text] [Medline: 26958264]
8. Ohashi K, Ota S, Ohno-Machado L, Tanaka H. Smart medical environment at the point of care: auto-tracking clinical interventions at the bed side using RFID technology. Comput Biol Med 2010 Jun;40(6):545-554. [doi: 10.1016/j.compbiomed.2010.03.007] [Medline: 20471637]
9. Halamka J. Early experiences with positive patient identification. J Healthc Inf Manag 2006;20(1):25-27. [Medline: 16429955]
10. Sundaresan S, Doss R, Zhou W. RFID in Healthcare – Current Trends and the Future. In: Adibi S, editor. Mobile Health. Cham, Switzerland: Springer; 2015:839-870.
11. Yao W, Chu CH, Li Z. The adoption and implementation of RFID technologies in healthcare: a literature review. J Med Syst 2012 Dec;36(6):3507-3525. [doi: 10.1007/s10916-011-9789-8] [Medline: 22009254]
12. Anand A, Fosso Wamba S. Business value of RFID‐enabled healthcare transformation projects. Bus Process Manag J 2013;19(1):111-145. [doi: 10.1108/14637151311294895]

XSL•FO
RenderX

13. Mun I, Kantrowitz AB, Carmel PW, Mason KP, Engels DW. Active RFID system augmented with 2D barcode for asset management in a hospital setting. In: Proceedings of the 2007 IEEE International Conference on RFID. 2007 Presented at: RFID '07; March 26-28, 2007; Grapevine, TX, USA p. 205-211. [doi: 10.1109/rfid.2007.346170]
14. Henry J, Pylypchuk Y, Searcy T, Patel V. ONC data brief: Adoption of electronic health record systems among US non-federal acute care hospitals: 2008–2015. The Office of the National Coordinator for Health Information Technology. 2016 May. URL: https://www.healthit.gov/sites/default/files/briefs/2015_hospital_adoption_db_v17.pdf [accessed 2022-05-04]
15. Jha AK, DesRoches CM, Campbell EG, Donelan K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. N Engl J Med 2009 Apr 16;360(16):1628-1638. [doi: 10.1056/NEJMsa0900592] [Medline: 19321858]
16. Angst CM, Agarwal R, Sambamurthy V, Kelley K. Social contagion and information technology diffusion: the adoption of electronic medical records in U.S. hospitals. Manag Sci 2010 Aug;56(8):1219-1241. [doi: 10.1287/mnsc.1100.1183]
17. Furukawa MF, Raghu TS, Spaulding TJ, Vinze A. Adoption of health information technology for medication safety in U.S. Hospitals, 2006. Health Aff (Millwood) 2008;27(3):865-875. [doi: 10.1377/hlthaff.27.3.865] [Medline: 18474981]
18. Yang X, Yang N, Lewis D, Parton J, Hudnall M. Patterns and influencing factors of eHealth tools adoption among Medicaid and non-Medicaid populations from the Health Information National Trends Survey (HINTS) 2017-2019: questionnaire study. J Med Internet Res 2021 Feb 18;23(2):e25809 [FREE Full text] [doi: 10.2196/25809] [Medline: 33599619]
19. Kazley AS, Ozcan YA. Organizational and environmental determinants of hospital EMR adoption: a national study. J Med Syst 2007 Oct;31(5):375-384. [doi: 10.1007/s10916-007-9079-7] [Medline: 17918691]
20. Bradley RV, Esper TL, In J, Lee KB, Bichescu BC, Byrd TA. The joint use of RFID and EDI: implications for hospital performance. Prod Oper Manag 2018 Nov 13;27(11):2071-2090. [doi: 10.1111/poms.12955]
21. Álvarez López Y, Franssen J, Álvarez Narciandi G, Pagnozzi J, González-Pinto Arrillaga I, Las-Heras Andrés F. RFID technology for management and tracking: e-health applications. Sensors (Basel) 2018 Aug 13;18(8):2663 [FREE Full text] [doi: 10.3390/s18082663] [Medline: 30104557]
22. Yazici HJ. An exploratory analysis of hospital perspectives on real time information requirements and perceived benefits of RFID technology for future adoption. Int J Inf Manag 2014 Oct;34(5):603-621. [doi: 10.1016/j.ijinfomgt.2014.04.010]
23. Bonkowski J, Carnes C, Melucci J, Mirtallo J, Prier B, Reichert E, et al. Effect of barcode-assisted medication administration on emergency department medication errors. Acad Emerg Med 2013 Aug;20(8):801-806 [FREE Full text] [doi: 10.1111/acem.12189] [Medline: 24033623]
24. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim WY, Chow A. Use of a real-time locating system for contact tracing of health care workers during the COVID-19 pandemic at an Infectious Disease Center in Singapore: validation study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
25. Galligioni E, Piras EM, Galvagni M, Eccher C, Caramatti S, Zanolli D, et al. Integrating mHealth in oncology: experience in the province of Trento. J Med Internet Res 2015 May 13;17(5):e114 [FREE Full text] [doi: 10.2196/jmir.3743] [Medline: 25972226]
26. Loveday A, Sherar LB, Sanders JP, Sanderson PW, Esliger DW. Technologies that assess the location of physical activity and sedentary behavior: a systematic review. J Med Internet Res 2015 Aug 05;17(8):e192 [FREE Full text] [doi: 10.2196/jmir.4761] [Medline: 26245157]
27. Huang Z, Guo H, Lee YM, Ho EC, Ang H, Chow A. Performance of digital contact tracing tools for COVID-19 response in Singapore: cross-sectional study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e23148 [FREE Full text] [doi: 10.2196/23148] [Medline: 33006944]
28. Dey A, Vijayaraman BS, Choi JH. RFID in US hospitals: an exploratory investigation of technology adoption. Manag Res Rev 2016 Apr 18;39(4):399-424. [doi: 10.1108/mrr-09-2014-0222]
29. Angst CM, Devaraj S, D'Arcy J. Dual role of IT-assisted communication in patient care: a validated structure-process-outcome framework. J Manag Inf Syst 2012;29(2):257-292. [doi: 10.2753/mis0742-1222290209]
30. Mahmood N, Shah A, Waqas A, Bhatti Z, Abubakar A, Malik HA. RFID based smart hospital management system: a conceptual framework. In: Proceedings of the 5th International Conference on Information and Communication Technology for The Muslim World. 2014 Presented at: ICT4M '14; November 17-18, 2014; Kuching, Malaysia p. 1-6. [doi: 10.1109/ICT4M.2014.7020594]
31. Guinard PF. Building a Smart Hospital using RFID technologies. In: Proceedings of the 2006 European Conference on e-Health. 2006 Presented at: ECEH '06; October 12-13, 2006; Fribourg, Switzerland p. 131-142.
32. Yao W, Chu CH, Li Z. Leveraging complex event processing for smart hospitals using RFID. J Netw Comput Appl 2011 May;34(3):799-810. [doi: 10.1016/j.jnca.2010.04.020]
33. Sharma N, Brinke JK, Van Gemert-Pijnen JE, Braakman-Jansen LM. Implementation of unobtrusive sensing systems for older adult care: scoping review. JMIR Aging 2021 Oct 06;4(4):e27862 [FREE Full text] [doi: 10.2196/27862] [Medline: 34612822]
34. Tai KY, Yen SL, Chen TS, Tseng YH. An implementation of interactive RFID positioning healthcare system on mobile platform. In: Proceedings of the 13th International Conference on Broadband and Wireless Computing, Communication and Applications. 2019 Presented at: BWCCA '18; October 27-29, 2018; Taichung, Taiwan p. 350-360. [doi: 10.1007/978-3-030-02613-4_31]

XSL•FO
RenderX

35. Xing Q, Miranda S, Singh R. Resource utilization uncertainty reduction in hospitals: the role of digital capability and governance structure. In: Proceedings of the 2020 International Conference on Information Systems. 2020 Presented at: ICIS '20; December 13-16, 2020; Hyderabad, India p. 12.
36. Carey K. Hospital cost efficiency and system membership. Inquiry 2003;40(1):25-38 [FREE Full text] [doi: 10.5034/inquiryjrnl_40.1.25] [Medline: 12836906]
37. Lin YK, Lin M, Chen H. Do electronic health records affect quality of care? Evidence from the HITECH act. Inf Syst Res 2019 Mar 12;30(1):306-318. [doi: 10.1287/isre.2018.0813]
38. Bardhan IR, Thouin MF. Health information technology and its impact on the quality and cost of healthcare delivery. Decis Support Syst 2013 May;55(2):438-449. [doi: 10.1016/j.dss.2012.10.003]
39. Appari A, Carian EK, Johnson ME, Anthony DL. Medication administration quality and health information technology: a national study of US hospitals. J Am Med Inform Assoc 2012;19(3):360-367 [FREE Full text] [doi: 10.1136/amiajnl-2011-000289] [Medline: 22037889]
40. Rosko MD, Proenca J, Zinn JS, Bazzoli GJ. Hospital inefficiency: what is the impact of membership in different types of systems? Inquiry 2007;44(3):335-349 [FREE Full text] [doi: 10.5034/inquiryjrnl_44.3.335] [Medline: 18038868]
41. Gross Domestic Product. Bureau of Economic Analysis. 2021. URL: https://www.bea.gov/data/gdp [accessed 2022-03-29]
42. Borzekowski R. Measuring the cost impact of hospital information systems: 1987-1994. J Health Econ 2009 Sep;28(5):938-949. [doi: 10.1016/j.jhealeco.2009.06.004] [Medline: 19699542]
43. Corrales J, Westhoff F. Information technology adoption and political regimes. Int Studies Q 2006 Dec;50(4):911-933. [doi: 10.1111/j.1468-2478.2006.00431.x]
44. Foster AD, Rosenzweig MR. Microeconomics of technology adoption. Annu Rev Econom 2010 Sep 01;2:10 [FREE Full text] [doi: 10.1146/annurev.economics.102308.124433] [Medline: 24386501]
45. Bazzoli GJ, Shortell SM, Dubbs N, Chan C, Kralovec P. A taxonomy of health networks and systems: bringing order out of chaos. Health Serv Res 1999 Feb;33(6):1683-1717 [FREE Full text] [Medline: 10029504]
46. Vlasakakis G, Comets E, Keunecke A, Gueorguieva I, Magni P, Terranova N, et al. White paper: landscape on technical and conceptual requirements and competence framework in drug/disease modeling and simulation. CPT Pharmacometrics Syst Pharmacol 2013 May 01;2(5):e40 [FREE Full text] [doi: 10.1038/psp.2013.16] [Medline: 23887723]
47. Burke DE, Wang BB, Wan TT, Diana ML. Exploring hospitals' adoption of information technology. J Med Syst 2002 Aug;26(4):349-355. [doi: 10.1023/a:1015872805768] [Medline: 12118818]
48. Wang SW, Chen WH, Ong CS, Liu L, Chuang YW. RFID application in hospitals: a case study on a demonstration RFID project in a Taiwan hospital. In: Proceedings of the 39th Annual Hawaii International Conference on System Sciences. 2006 Presented at: HICSS '06; January 4-7, 2006; Kauai, HI, USA p. 184a. [doi: 10.1109/HICSS.2006.422]
49. Lin SC, Jha AK, Adler-Milstein J. Electronic health records associated with lower hospital mortality after systems have time to mature. Health Aff (Millwood) 2018 Jul;37(7):1128-1135. [doi: 10.1377/hlthaff.2017.1658] [Medline: 29985687]
50. Lee J, McCullough JS, Town RJ. The impact of health information technology on hospital productivity. RAND J Econ 2013;44(3):545-568. [doi: 10.1111/1756-2171.12030]
51. Jha AK, DesRoches CM, Shields AE, Miralles PD, Zheng J, Rosenbaum S, et al. Evidence of an emerging digital divide among hospitals that care for the poor. Health Aff (Millwood) 2009;28(6):w1160-w1170. [doi: 10.1377/hlthaff.28.6.w1160] [Medline: 19858142]
52. AHA Annual Survey Database. American Hospital Association. URL: https://www.ahadata.com/aha-annual-survey-database [accessed 2022-05-04]

Abbreviations

AHA: American Hospital Association
GDP: gross domestic product
HIT: health information technology
IT: information technology
OR: odds ratio
RFID: radio-frequency identification
VPC: visual predictive check

XSL•FO
RenderX

Edited by T Leung; submitted 21.09.21; peer-reviewed by H Hah; comments to author 13.10.21; revised version received 14.11.21; accepted 14.04.22; published 26.05.22.

Please cite as:
Zhu X, Tao Y, Zhu R, Wu D, Ming WK
Impact of Hospital Characteristics and Governance Structure on the Adoption of Tracking Technologies for Clinical and Supply Chain Use: Longitudinal Study of US Hospitals
J Med Internet Res 2022;24(5):e33742
URL: https://www.jmir.org/2022/5/e33742
doi:10.2196/33742
PMID:35617002

©Xiao Zhu, Youyou Tao, Ruilin Zhu, Dezhi Wu, Wai-kit Ming. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2022. 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.

XSL•FO
RenderX

Original Paper

A Virtual Community of Practice to Support Physician Uptake of a Novel Abortion Practice: Mixed Methods Case Study

Sheila Dunn[1,2], MSc, MD; Sarah Munro[3,4], PhD; Courtney Devane[5], RN, MN; Edith Guilbert[6], MSc, MD; Dahn Jeong[7], MSc; Eleni Stroulia[8], PhD; Judith A Soon[9], RPh, PhD; Wendy V Norman[10,11], MHSc, MD

[1]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[2]Women's College Research Institute, Women's College Hospital, Toronto, ON, Canada
[3]Department of Obstetrics and Gynaecology, University of British Columbia, Vancouver, BC, Canada
[4]Centre for Health Evaluation and Outcome Science, University of British Columbia, Vancouver, BC, Canada
[5]School of Nursing, Faculty of Applied Science, University of British Columbia, Vancouver, BC, Canada
[6]Department of Obstetrics, Gynecology and Reproduction, Laval University, Quebec City, QC, Canada
[7]School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[8]Department of Computer Science, Faculty of Science, University of Alberta, Edmonton, AB, Canada
[9]Contraception and Abortion Research Team, Women's Health Research Institute, Provincial Health Services Authority, Vancouver, BC, Canada
[10]Department of Family Practice, University of British Columbia, Vancouver, BC, Canada
[11]Faculty of Public Health & Policy, London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Sheila Dunn, MSc, MD
Department of Family and Community Medicine
University of Toronto
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400 ext 4977
Fax: 1 416 323 6351
Email: sheila.dunn@wchospital.ca

Abstract

Background: Virtual communities of practice (VCoPs) have been used to support innovation and quality in clinical care. The drug mifepristone was introduced in Canada in 2017 for medical abortion. We created a VCoP to support implementation of mifepristone abortion practice across Canada.

Objective: The aim of this study was to describe the development and use of the Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement (CAPS-CPCA) VCoP and explore physicians' experience with CAPS-CPCA and their views on its value in supporting implementation.

Methods: This was a mixed methods intrinsic case study of Canadian health care providers' use and physicians' perceptions of the CAPS-CPCA VCoP during the first 2 years of a novel practice. We sampled both physicians who joined the CAPS-CPCA VCoP and those who were interested in providing the novel practice but did not join the VCoP. We designed the VCoP features to address known and discovered barriers to implementation of medication abortion in primary care. Our secure web-based platform allowed asynchronous access to information, practice resources, clinical support, discussion forums, and email notices. We collected data from the platform and through surveys of physician members as well as interviews with physician members and nonmembers. We analyzed descriptive statistics for website metrics, physicians' characteristics and practices, and their use of the VCoP. We used qualitative methods to explore the physicians' experiences and perceptions of the VCoP.

Results: From January 1, 2017, to June 30, 2019, a total of 430 physicians representing all provinces and territories in Canada joined the VCoP and 222 (51.6%) completed a baseline survey. Of these 222 respondents, 156 (70.3%) were family physicians, 170 (80.2%) were women, and 78 (35.1%) had no prior abortion experience. In a survey conducted 12 months after baseline, 77.9% (120/154) of the respondents stated that they had provided mifepristone abortion and 33.9% (43/127) said the VCoP had been important or very important. Logging in to the site was burdensome for some, but members valued downloadable resources such as patient information sheets, consent forms, and clinical checklists. They found email announcements helpful for keeping

XSL•FO
RenderX

up to date with changing regulations. Few asked clinical questions to the VCoP experts, but physicians felt that this feature was important for isolated or rural providers. Information collected through member polls about health system barriers to implementation was used in the project's knowledge translation activities with policy makers to mitigate these barriers.

Conclusions: A VCoP developed to address known and discovered barriers to uptake of a novel medication abortion method engaged physicians from across Canada and supported some, including those with no prior abortion experience, to implement this practice.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2018-028443

(J Med Internet Res 2022;24(5):e34302) doi:10.2196/34302

KEYWORDS

mifepristone; abortion; community of practice; virtual community of practice; diffusion of innovation; learning community

Introduction

Background

Communities of practice (CoPs) are recognized as tools for enhancing knowledge, improving practice, and supporting innovation [1,2]. As described by Wenger et al [3], CoPs are "groups of people who share a concern, a set of problems or a passion about a topic and who deepen their knowledge and expertise in this area by interacting on an ongoing basis." In health care, CoPs have been used to exchange information and knowledge, support implementation of practice innovations, build a sense of identity, and reduce professional isolation [1,4-6]. Virtual CoPs (VCoPs) can achieve these goals among geographically dispersed practitioners [5,7]. We hypothesized that a VCoP could be particularly valuable to promote the adoption of a novel medical practice (mifepristone abortion care) introduced in Canada in 2017 and to facilitate uptake of this practice by primary care providers, particularly those in rural and remote regions who may have limited professional support and resources [7-9].

Mifepristone, when used in combination with misoprostol, is recognized internationally as the *gold standard* for medication abortion [10]. Since its first approval in France in 1988, mifepristone has been approved in more than 79 countries and has been used by millions of people worldwide [11]. Mifepristone is safe, effective, and as a straightforward alternative to surgical abortion has transformed the way abortion is provided; its introduction in Canada in 2017 raised the important question of how health care professionals could be helped to implement this innovation.

In preparation for mifepristone's availability in 2017, the Contraception and Abortion Research Team-Groupe de recherche sur l'avortement et la contraception [12] launched the CART-Mife Study, a 4-year national implementation research project, described elsewhere [13], that aimed to identify and mitigate barriers to implementing mifepristone abortion practice, particularly those affecting community-based physicians and pharmacists. The project had two interventions: (1) integrated knowledge translation with health policy makers to mitigate health system barriers and support facilitators to adoption of mifepristone abortion practice by physicians and pharmacists and (2) a VCoP [14], which was established to address the needs of community-based physicians and pharmacists across Canada who were interested in adopting mifepristone abortion practice in their professional roles.

Canada's laws, regulations, geography, and health system present challenges and opportunities for mifepristone abortion practice that are distinct from those in other countries. Almost unique in the world, Canada has no criminal law on abortion [15]. Since 1988, abortion has been considered a medical procedure and its need determined by the patient with their health care provider [16]. In addition to imposing varied criminal sanctions on abortion, most high-income countries also highly regulate how mifepristone is prescribed and dispensed, as well as where it is used [17-19]. Mifepristone's initial Canadian approval in 2015 had several similar regulatory restrictions, most of which were removed over the 2 years after its availability in 2017 [20,21]. Currently, many countries restrict mifepristone provision to certain types of practitioners, such as medical specialists or registered approved providers, or purpose-specific facilities [22-24]. Except for the province of Quebec [25], Canada has eliminated such restrictions and allows prescription by any authorized prescriber (physicians and, in most provinces, also nurse practitioners) [26]. Some countries, including the United States, have not allowed pharmacies to dispense mifepristone but require drug dispensing by the prescriber or clinic [17,24]. Not so in Canada; by November 2017, mifepristone became available from pharmacies like any other drug, to be dispensed by any pharmacist when presented with a prescription [26]. Government health insurance plans cover costs of the drug. The requirement for preabortion ultrasound was removed in April 2019 [27]. Although ultrasound is often used, clinical guidelines and mifepristone drug approvals in most other countries, including the United States, do not require it [10,17,19]. Canada's lack of restrictions opened the door for mifepristone abortion provision in community primary care. This globally unique situation presented an opportunity to address the considerable geographical challenges to abortion access faced by Canadians living in rural or remote communities, distant from the large metropolitan centers where most abortion services are located [28]. Availability in community primary care could provide patients with local access to abortion through their own health care provider.

To promote widespread uptake of mifepristone abortion practice, addressing nonregulatory barriers to implementation was also crucial. Medication abortion care is not complicated or difficult, but in 2017 few Canadian physicians were knowledgeable about, or had experience with, it. Abortion is also highly stigmatized.

XSL•FO
RenderX

The fear of negative attitudes or harassment from colleagues, patients, and local communities could inhibit and isolate abortion providers [29-32]. In 2017, Dawson et al [33] identified challenges and facilitators experienced by primary care medication abortion providers in Australia, whose publicly funded health system and mifepristone drug approval are similar to those in Canada. Barriers included not recognizing medication abortion as within the physician's scope of practice; stigma; logistical challenges such as finding a pharmacy with the drug, access to ultrasound, consent forms, and patient information sheets; lack of experience, access to experts, mentorship, and peer support; and professional isolation. In Canada, Dressler et al [8] found that rural physicians also experienced professional isolation and lack of training opportunities.

We theorized that information, resources, tools for practice, and an accessible professional network to access and share implementation enablers would enhance the ability, and perhaps willingness, of approved health care providers (initially only physicians and pharmacists) to provide medication abortion care. Furthermore, we theorized that real-time collection of reported barriers could inform health system and regulatory decision-makers' understanding and ability to address unanticipated barriers. Working with national health professional organizations, guideline committees, and government regulators, we developed a national VCoP with these features. Our VCoP went live in January 2017, at the same time that mifepristone became commercially available.

Objectives

This paper describes the development of the Canadian Abortion Providers Support-Communauté de pratique Canadienne sur l'avortement (CAPS-CPCA) VCoP, examines its use, and explores the perspectives of physicians, who were the only eligible prescribers at the start of the study period, on its value for implementing this novel clinical practice.

Methods

We adopted an intrinsic case study approach using mixed methods during the study period January 1, 2017, to October 30, 2019. Intrinsic case studies are used to explore a specific event or issue in depth in a real-life context [34].

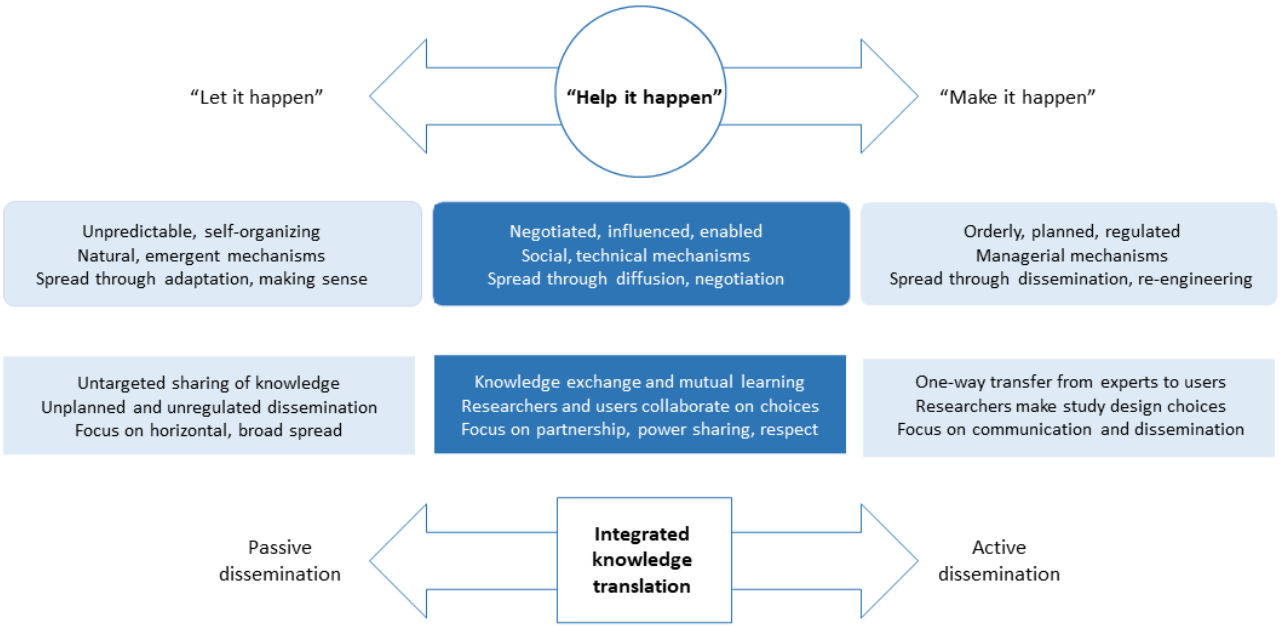
Theoretical Framework for CAPS-CPCA

Our development of CAPS-CPCA was informed by the *Theory of the Diffusion of Innovation* formulated by Rogers [35] as operationalized by Greenhalgh et al [36]. Greenhalgh et al [36] theorized that the implementation of innovations in health systems is affected by a complex interaction of influences. These include characteristics of the innovation (complexity, compatibility, advantage, trialability, and observability) and the adopter (motivation, skills, and values), system readiness for the innovation (tension for change, innovation-system fit, and dedicated resources), mechanisms used for implementation (technical support and social networks), the outer context of regulatory and sociopolitical influences, and the communication and influence of change agents and knowledge purveyors. We conceptualized the CAPS-CPCA VCoP as a mechanism to support this innovation in abortion practice—in the words of Greenhalgh et al [36], to "Help it happen"—through both social and technical means [37] (Figure 1 [36-38]).

We anticipated that most VCoP members would share common attitudes (homophily) and be motivated to join because of an interest in, and commitment to, women's reproductive health that included abortion. The VCoP features aimed to decrease complexity and increase compatibility of the innovation, explain and improve the innovation's relative advantage for adopters, reduce its perceived risks, and create a social network to enhance knowledge and share experience and expertise. We also used the CAPS-CPCA VCoP as a tool to identify physician and pharmacist experiences of health policy and systems barriers to implementation—findings that informed the main project's integrated knowledge translation activities with health policy decision-makers to mitigate or eliminate these barriers (the regulatory and sociopolitical influences described by Greenhalgh et al [36]), and further help these practitioners to adopt this practice [13,37]. Integrated knowledge translation is the process of including key stakeholders in all stages of the research process, which in our study included discussing the barriers and facilitators data collected through the CAPS-CPCA VCoP with federal decision-makers in real time.

XSL•FO
RenderX

Figure 1. Conceptual framework for diffusion of innovation (reproduced from Munro et al [37], which is published under Creative Commons Attribution 4.0 International License [39].



From Munro et al [37]. Adapted from concepts by Greenhalgh et al [36] and Bowen and Graham [38].

CoP Development

The CAPS-CPCA web-based platform (Multimedia Appendix 1) allowed members asynchronous access through their personal account to site content (in English and French), to find pharmacies in their community dispensing mifepristone, to post comments and tips, or ask questions. CAPS-CPCA aimed to encourage multidirectional interaction among members, experts, and researchers to promote sharing information of best practice resources and practice facilitators. Its features specifically addressed barriers to medication abortion practice that had been identified in the literature or were identified during the research project [8,33] (Table 1).

Table 1. Features of the virtual community of practice addressing barriers and facilitators to mifepristone abortion uptake and related Diffusion of Innovation constructs.

Barrier or facilitator (Diffusion of Innovation constructs addressed)	CAPS-CPCA[a] feature
Lack of clinical knowledge (advantage, complexity, experience, risk, and observability)	• Clinical practice guidelines and reviews • Frequently Asked Questions • Email Member Announcements
Tools for practice (complexity and compatibility)	• Sample forms (consent, patient information, and follow-up forms, as well as charting forms and checklists) • Patient resources • Billing codes
Logistical challenges (social values, trialability, diffusion and influence, and system readiness)	• Discussion Room • Map of pharmacies stocking mifepristone • "What's happening in your province?" • Member polls
Peer support and access to experts (social values, trialability, diffusion and influence, and system readiness)	• Discussion Room • Ask an Expert
Isolation and stigma (social values, trialability, diffusion and influence, and system readiness)	• Membership in CAPS-CPCA • Discussion Room • Ask an Expert • Email Member Announcements • Member polls

[a]CAPS-CPCA: Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement.

To address the desire for peer support, expert advice, and mentorship, CAPS-CPCA had participative *Discussion Room* and *Ask an Expert* features. Members could pose clinical or practice-related questions and receive a response from an

XSL•FO
RenderX

experienced Canadian abortion provider within 24 hours. To maintain confidentiality, questions and answers were not directly visible to all site users and all posted interactions were identified only by the user's ID: a random number tagged with a professional 2-letter designation (eg, 3174MD and 2061NP). Reoccurring clinical questions were rephrased and shared with all users through *Member Announcements* and *Frequently Asked Questions* features.

We sought formal support from professional organizations representing most of the anticipated providers from across Canada (ie, family medicine, obstetrics and gynecology, nursing, and pharmacy) for their social influence among their members to confirm that mifepristone abortion was within their members' scope of practice and build credibility of the VCoP. We included their organizational logos in branding materials. Research team members were recognized local and national experts in their disciplines and promoted CAPS-CPCA through their professional networks. To further inspire confidence and minimize perceived conflict of interest, we received no industry funding and did not disseminate industry-prepared materials.

Throughout the project we responded to member requests for additional support; for example, we created resources such as clinical checklists and guideline summaries [40]. Acting on early feedback that many members preferred email notifications rather than logging in to the website, we began emailing *Member Announcements* containing brief information on policy updates, continuing education events, common practice questions, relevant research, and product shortages. Finally, email polls allowed members to contribute information about the real-time impact of health policies, such as the early requirement for prescribers to register with the manufacturer or for a mandatory ultrasound to initiate medication abortion, knowing that the member's perspective would be used to inform policy decisions.

CAPS-CPCA Member Recruitment

Vigorous recruitment was a key strategy for community building to reduce isolation and stigma associated with abortion practice. We invited interested physicians and pharmacists from across Canada to join CAPS-CPCA. This recruitment occurred primarily through a web-based medical abortion training course hosted by the Society of Obstetricians and Gynaecologists of Canada that, until November 2017, was a prerequisite to prescribe or dispense mifepristone [41]. Other recruitment occurred through continuing education events, word of mouth, announcements from our partner organizations to their members, at the federal drug regulator's drug information site and in its communications distributed to all practicing physicians, and on the product website. After the removal of Health Canada's regulation for physician-only prescribing in November 2017, CAPS-CPCA extended its membership to nurse practitioners and midwives through their professional organizations.

Acknowledging concerns about safety and potential for harassment, membership was restricted to licensed health professionals (physicians, pharmacists, and later nurse practitioners and midwives), their verified staff, health professional trainees, and project collaborators. Internal firewalls permitted only the licensed health professionals to access the site's clinical discussion and expert advice areas. Membership requests were made on the web and vetted by the research team by verifying the applicant's professional license and requesting references, if needed.

Data Sources

We collected data from three sources during the study period:

1. CAPS-CPCA website data and *Member Announcements* WordPress data collected from January 1, 2017, to June 30, 2019, included member details, page views or downloads accessible only to members, resource views or downloads accessible from the landing page, *Member Announcements* emails opened, and email poll responses. We also collected content from *Ask an Expert* questions and *Discussion Room* threads with physician posts.
2. Electronic surveys were completed as part of the main CART-Mife Study by CAPS-CPCA physician members and nonmembers who were interested in providing medication abortion. Survey development is described elsewhere [42]. Surveys were administered in English or French at baseline, 6 months, and 12 months (last 12-month survey collected in October 2019) to collect data on clinician characteristics and practices as well as barriers and facilitators to implementation (Multimedia Appendices 2 and 3). Follow-up surveys included 7 questions about CAPS-CPCA participation, its importance, and suggestions for improvement.
3. Interviews: As part of the main study, we conducted semistructured interviews in English or French with a national sample of abortion-providing and nonproviding physicians, including a subset of survey respondents, as well as health system stakeholders. Details of the qualitative study design and results of the interviews are reported elsewhere [13,21,25,37,43]. Interviews were conducted by telephone in English or French by a knowledge translation scientist (SM), physician researcher (EG), and nursing doctoral student (CD). The interview questions probed for domains of the Diffusion of Innovation theory. Of relevance to CAPS-CPCA, specific questions explored VCoP membership, how it did or did not support prescribers, and their experience and overall thoughts about the VCoP. For participants who had not accessed CAPS-CPCA, we asked if joining this website would be useful (why or why not), what information they would want from the website, and what features they liked about other CoPs.

Although all CAPS-CPCA members representing diverse health professions (physicians, pharmacists, nurses and nurse practitioners, and midwives) are included in the overall site use data, the data from surveys, interviews, and website *Ask an Expert* and *Discussion Room* content used in this analysis relate only to physicians, who were the only health care providers initially eligible to prescribe mifepristone and who made up most (430/521, 82.5%) of the eligible prescribers throughout the study.

Data Analysis

Site metrics, CAPS-CPCA member characteristics, and responses to survey questions were analyzed descriptively (counts, means, medians, and percentages), and we used

XSL•FO
RenderX

chi-square statistics to examine the association of member characteristics with members' reported use and perceived importance of the VCoP with significance set at $P<.05$. Website page views were aggregated from the webserver logs using AWStats [44]. We analyzed qualitative data (website threads, open-ended survey responses, and interviews) drawing from directed content analysis and thematic analysis approaches, using concepts from the Diffusion of Innovation theory as guiding deductive codes, which we then tested and refined with inductive coding [35,45-47]. We examined and categorized content from *Ask an Expert* and *DiscussionRoom* threads as related to system and regulatory, implementation and logistical, or clinical issues. We analyzed audio-recorded, transcribed interviews for themes related to our key objectives for this substudy and explored physician participants' use of the CAPS-CPCA and their perspectives on the value of the VCoP for implementing this clinical practice. Methods of thematic analysis and additional results of our analysis of the interviews are described in previous publications [21,25,43]. Using mixed methods techniques, we triangulated our data concurrently with individual data set analyses to compare and contrast findings and gain a deeper understanding of how members used the CAPS-CPCA VCoP and why [48].

Ethics Approval

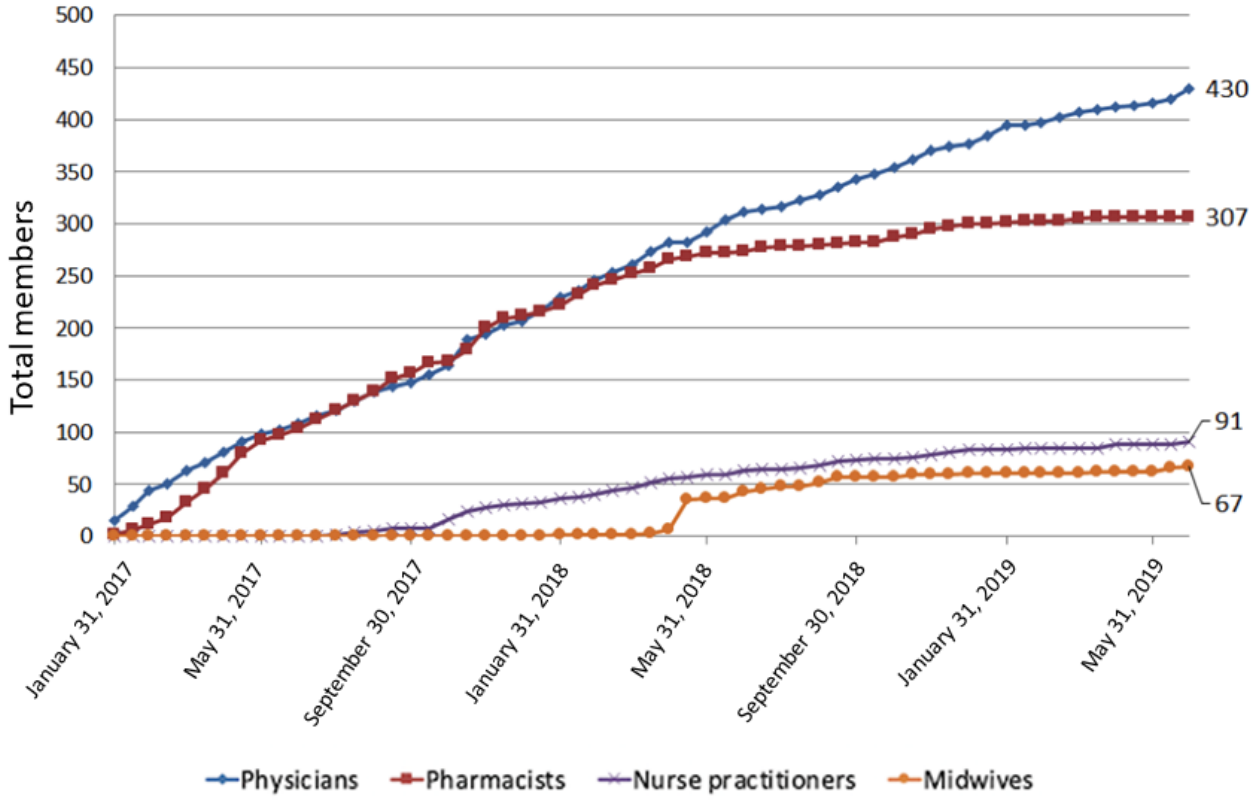
The CART-Mife Study received ethics approval from the Behavioural Research Ethics Board at the University of British Columbia (H16-01006).

Results

Overview

Over the first 30 months, CAPS-CPCA membership grew steadily, accepting more than 1000 members representing all provinces and territories, including 430 physicians (Figure 2). Of the 430 CAPS-CPCA physicians, 222 (51.6%) participated in the baseline survey available from January 2017 until April 2019, which collected demographics and abortion experience (Table 2). Of the 222 respondents, 170 (80.2%) were female; 156 (70.3%) were family physicians; 15 (6.8%) practiced in regions with no abortion services before January 2017; 78 (35.1%) had no previous abortion experience; and, notably, 123 (55.4%) practiced outside metropolitan areas, although only 29.5% of the Canadian population live there [49].

Figure 2. Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement clinician membership over time.



XSL•FO
RenderX

Table 2. CAPS-CPCA[a] virtual community of practice: characteristics of physician members who completed a baseline survey (N=222).

Characteristic	Values, n (%)
Sex	
Female	170 (80.2)
Male	38 (17.1)
Other or missing	6 (2.7)
Age (years)	
<35	60 (27)
35 to 44	56 (25.2)
45 to 54	53 (23.9)
>54	47 (21.2)
Missing	6 (2.7)
Primary specialty	
Family or general practice	156 (70.3)
Obstetrician-gynecologist	53 (23.9)
Medical student or resident	10 (4.5)
Other or missing	3 (1.4)
Province (% of Canadian population) [50]	
Ontario (38.6)	82 (36.9)
British Columbia (13.5)	39 (17.6)
Quebec (22.6)	30 (13.5)
Nova Scotia (2.6)	17 (7.7)
Saskatchewan (3.1)	13 (5.9)
Alberta (11.6)	12 (5.4)
Manitoba (3.6)	6 (2.7)
Atlantic provinces[b] (3.9)	10 (4.5)
Northern territories[c] (0.3)	9 (4.1)
Missing	4 (1.8)
Residence location (% of Canadian population) [49]	
Large metropolitan area (71.8)	94 (42.3)
Outside large metropolitan area (29.5)	123 (55.4)
Missing	5 (2.3)
Previous abortion experience	
None	78 (35.1)
Medical and surgical	71 (32)
Medical only	36 (16.2)
Surgical only	33 (14.9)
Missing	4 (1.8)
Primary facility type	
Private physician office	78 (35.2)
Community abortion or reproductive health clinic	40 (18)
General health care community or ambulatory clinic	25 (11.3)
Hospital-affiliated facility	54 (24.3)

XSL•FO
RenderX

Characteristic	Values, n (%)
Other	5 (2.3)
Missing	20 (9)
Other abortion services available in the community	
Medical and surgical	134 (60.4)
Surgical only	33 (14.9)
Medical only	20 (9)
None	15 (6.8)
Missing	20 (9)
Do you currently, or do you plan to, prescribe mifepristone?	
Yes	144 (64.9)
No	16 (7.2)
Missing	62 (27.9)

[a]CAPS-CPCA: Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement.

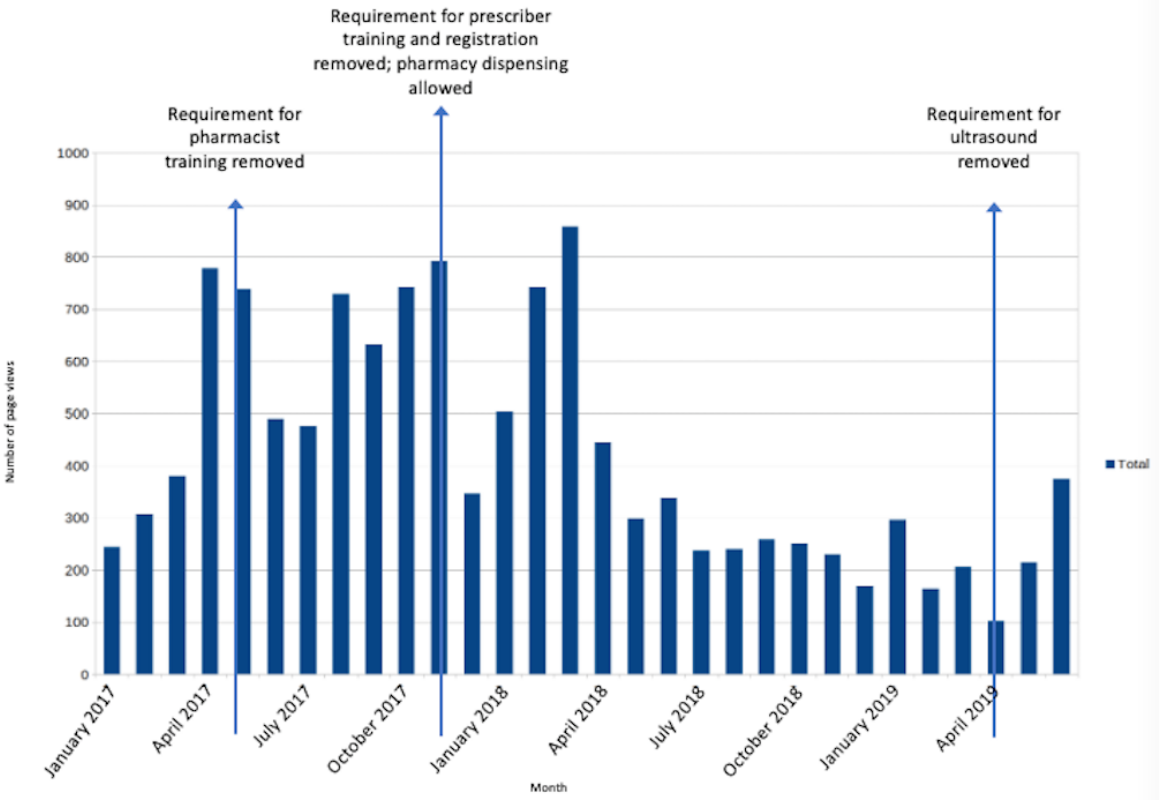
[b]New Brunswick, Newfoundland and Labrador, and Prince Edward Island were combined because of small cell sizes.

[c]Yukon, Northwest Territories, and Nunavut were combined because of small cell sizes.

Website Data

Website traffic is shown in Figure 3. Traffic peaked in the first half of the study period and then declined but remained stable. The most frequently visited pages were *Helpful Resources* (2338/12,592, 18.57%, page visits), *Locate a Pharmacy* (2154/12,592, 17.11%), *Ask an Expert* (1792/12,592, 14.23%), and *Latest News* (1892/12,592, 15.03%). From January 1, 2017, to June 30, 2019, there were more than 10,000 views or downloads of resources, including some (ie, *Prescriber* and *Pharmacist Checklists* and *Prescriber* and *Pharmacist Resource Guides*) which, at the request of Health Canada, were made openly available on the CAPS-CPCA landing page and thus were not exclusive to members. The *Prescriber Resource Guide* and *Prescriber Checklist* were viewed or downloaded 1759 times. Other resources accessible only to VCoP members and most relevant to prescribers, such as consent forms, patient information sheets, pharmacy locations, and information on coverage for drug costs across the country, had 1263 views or downloads.

Figure 3. Page views from January 1, 2017, to June 30, 2019, with regulatory-change dates.



XSL•FO
RenderX

Each of the 77 email *Member Announcements* was opened an average of 341.8 (SD 73.3) times. Response to 2 email polls distributed on May 30, 2017, and March 22, 2019, was 48% (47/99) of the physicians and 5.7% (28/489) of the prescribers (physicians and nurse practitioners), respectively. The first poll asked about the early requirement for physicians to register with the pharmaceutical company to prescribe and dispense mifepristone, similar to a current requirement in the United States [17]. The respondents did not support this restrictive requirement, which also violated codes of conduct of some provincial licensing bodies [51]. It was removed the same week that the poll results were shared with the regulatory decision-makers. The second poll asked about the requirement for pelvic ultrasound in the initial drug approval [20]. Of the 28 respondents, 9 (32%) said that mandatory ultrasound limited their ability to provide mifepristone abortion. Health Canada subsequently removed this requirement [27].

During the 30-month website data collection period, physicians posed 38 *Ask an Expert* questions and there were 19 physician posts in 14 *Discussion Room* threads. Of the 52 questions and discussions, 12 (23%) related to health system or regulatory issues such as how to access mifepristone through pharmacies and hospitals, access to surgical abortion for failed abortions, billing for medication abortion, and drug shortages; 12 (23%) related to logistics of implementation, such as on-call coverage, considerations in rural and remote areas, and overcoming resistance of colleagues or hospitals to mifepristone abortion; and 28 (54%) were clinical questions ranging from use of mifepristone in specific circumstances (eg, breastfeeding, breast cancer, and opiate user), more complicated clinical courses (eg, lack of bleeding or persistent bleeding), and recommended practice when clinical or laboratory resources were limited (eg, access to ultrasound and management of Rh-negative patients).

Survey Data

CAPS-CPCA members constituted 56.3% (129/229) of the respondents in the 6-month physician follow-up survey, with 66.7% (86/129) of these member respondents agreeing or strongly agreeing that the VCoP was helpful (data not shown). The 12-month follow-up surveys were completed by 224 physicians, of whom 127 (56.7%) indicated that they were still CAPS-CPCA members (Table 3). Of these 127 members, 81 (63.8%) said that the availability of a web-based support platform had been important or very important to them when deciding to provide mifepristone. Perceived importance of the VCoP was not associated with physician experience or urban or rural location. At 12 months, 59.1% (75/127) of the physicians intended to remain members of CAPS-CPCA. In open-ended responses, many members commented that logging in to the site was burdensome and that site navigation should be improved but they valued email updates and the resources and tools for practice. Several remarked that for clinical concerns, they preferred local professional contacts rather than CAPS-CPCA peers or experts because “CAPS...is less helpful for in-the-moment clinical support.” Others recommended that CAPS-CPCA build a centralized resource to support referrals for needed clinical backup and a larger list of pharmacies stocking the medication. Of the 57 respondents who said that they were not members, 34 (58%) had not heard of CAPS-CPCA and those who *were* aware of the VCoP cited no need for it and lack of time as the main reasons for not joining.

XSL•FO
RenderX

Table 3. Members' opinions about Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement (CAPS-CPCA): 12-month–survey responses.

	Values, n (%)
Respondents to 12-month survey (N=224)	
CAPS-CPCA members	127 (56.7)
Nonmembers	57 (25.4)
Missing	40 (17.9)
How important was it to know there was an online platform for support when you decided to provide mifepristone? (n=127)[a]	
Important or very important	81 (63.8)
Neutral or not important	46 (36.2)
Now (after 1 year), how important has the CoP[b] been? (n=127)[a]	
Important or very important	43 (33.8)
Neutral or not important	70 (55.1)
Missing	14 (11)
Do you plan to continue to participate in CAPS-CPCA? (n=127)[a]	
Yes	75 (59.1)
No	13 (10.2)
Don't know	39 (30.7)
How many times in the past 12 months have you accessed CAPS-CPCA? (n=127)[a]	
0	34 (26.8)
1 to 2	44 (34.6)
3 to 5	25 (19.7)
>5	18 (14.2)
Missing	6 (4.7)

[a]Only those stating that they were members of Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement were asked questions related to the virtual community of practice.

[b]COP: community of practice.

By 12 months, 77.9% (120/154) of the respondents who had ever been a CAPS-CPCA member, including 65% (34/52) with no previous abortion experience, indicated that they had provided mifepristone abortions (median 20, IQR 3-50).

Interview Data

Over the first year of CAPS-CPCA, we conducted interviews with 55 physicians as part of our broader study on implementation of mifepristone in Canada [13,21], including 33 (60%) members and 22 (40%) nonmembers, 91% (20/22) of whom had not heard of CAPS-CPCA. Analysis identified the following key themes about CAPS-CPCA among members: sense of community and support, clinical usefulness of practice tools, the importance of access to clinical support, importance of CAPS-CPCA for keeping up to date on regulatory changes, preference for emails for information, and concerns about security (Textbox 1). Some interviewees were unfamiliar, or had not engaged, with CAPS-CPCA but felt that access to experts, practice tools, and information on regulations would be valuable in a VCoP.

XSL•FO
RenderX

Textbox 1. Thematic analysis of physician interviews.

Sense of community and support

- Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement (CAPS-CPCA) gave members a sense of community and support, often described as a sense of a community of practitioners, spread nationwide. A participant stated, "Well, it's just that you don't feel so alone" [Physician 013], and another said, "We need to have platforms that bring providers together to deal with whatever issues are arising" [Physician 018]. Another commented, "I also sort of just like the feeling that it makes you part of a community with people with common interests" [Physician 015]. Some felt that it was especially important for new providers and could increase their confidence to adopt the practice: "If you were a new provider going to it you would get the guidelines...the checklists...support, if you needed it" [Physician 002]; "A great resource to even tell people that are thinking about doing this and are feeling a little less confident" [Physician 019].

Clinical usefulness of practice tools

- Members and nonmembers agreed about the usefulness of downloadable practice support tools such as consent forms, guidelines, and checklists, as well as a more extensive list of dispensing pharmacies. Participants who acted as informal mentors described CAPS-CPCA as their go-to resource for educating new mifepristone providers and linking them with practice tools. In turn, participants provided suggestions on useful practice tools, which informed how we organized and shared resources on the platform: "If anything, I would say have more handouts that you could print off and give to patients...it would be nice to go to that abortion providers' website and just go, 'I know where I can find it,' because exactly. Sometimes, you're scrambling" [Physician 033].

Importance of clinical support

- Members and nonmembers articulated a need for access to clinical support. Some had established and preferred local contacts: "I would feel more comfortable just phoning up the obstetricians I have a close relationship with than to post something on a board and have people that I don't have [a] relationship with answer" [Physician 025]. However, they also felt that "[CAPS-CPCA is] an excellent resource and community...[to] go get support" [Physician 006], particularly for solo or rural physicians without a local support network: "I think for people that may be more kind of solo or in a group of family doctors, that may be a really helpful place to ask opinions on kind of situations you might come across" [Physician 004]. A participant expressed a desire for personalized mentorship: "It would also be good to have almost like a mentor to just touch base with every once in a while, potentially also to discuss more difficult cases" [Physician 042], and another wanted more local or regional subgroups for direct communication among members.

Importance of CAPS-CPCA for keeping up to date with information and regulatory changes

- "I've no idea how I would learn about [changing regulations] though, if I weren't on CAPS-CPCA" [Physician 016]. In early interviews, CAPS-CPCA members indicated a preference for emails for information and "tend[ed] to go less on the website" [Physician 002] because logging in was burdensome. However, some found the emails too frequent and even "intrusive" [Physician 013].

Concerns about security

- Physicians accepted that security issues were part of abortion care and stressed the importance of website security in that "people are pretty cautious about sort of publicly being identified as abortion providers" [Physician 035]. Members appreciated the fact that the website was run by known leaders as well as the process for member authentication. However, concerns about "databases where my name and info are potentially breachable" kept a physician from joining [Physician 035]. Physicians were skeptical about pharmaceutical industry involvement in clinical practice and a member voiced concern that CAPS-CPCA might be a "vehicle for promotion" for pharmaceutical companies [Physician 013].

Lack of awareness and engagement with the virtual community of practice

- Some interview participants were unfamiliar with CAPS-CPCA, notably French-speaking physicians from Quebec. Interviewees who were aware of the virtual community of practice and had elected not to join did not perceive it as useful to them; for example, among highly experienced abortion providers those not ready to implement mifepristone abortion practice, those who felt no need for the virtual community of practice, and those who perceived that they had no time for it. A physician who had integrated mifepristone protocols and materials into their electronic medical record felt "like there's nothing that a support group would help me with" [Physician 025].

Discussion

Principal Findings

In our intrinsic case study, we describe the development and use of a VCoP for mifepristone abortion providers during the first 30 months of its availability in Canada—a jurisdiction free of the legal and regulatory restraints present in many countries [15,17-19]. The alignment of findings from our website, surveys, and interviews demonstrates that CAPS-CPCA provided important support for some physicians wanting to implement this new practice. Our recruitment of 430 physician VCoP members from all regions of the country shows that many potential new medication abortion providers wanted support when mifepristone was introduced. Although we do not know what proportion of Canadian abortion providers these 430 physicians represent, this number is sizable. A 2012 study found fewer than 300 physicians across Canada who were providing abortions, most of them surgical [28]. Of our 222 CAPS-CPCA survey respondents, 140 (63.1%) had previous abortion experience, 111 (50%) had no experience with medication abortion, and 78 (35.1%) were new providers who had no experience with any type of abortion care. These findings suggest that the number of abortion providers is increasing, and emerging evidence supports this [52]. More than half (123/222, 55.4%) of the CAPS-CPCA survey respondents were from areas outside the large metropolitan centers where abortion services in Canada are concentrated [28]. New abortion providers,

XSL•FO
RenderX

working in areas where services are lacking, could increase equity of access to abortion for Canadians.

Recruitment to CAPS-CPCA was very low in Quebec, and most Quebec physicians interviewed were unaware of it. System readiness for this innovation was low and inflexible in Quebec. The College of Physicians of Quebec placed explicit restrictions on the conditions for prescribing mifepristone, and physicians perceived administrative complexities to implementing medication abortion protocols. There was also a noted resistance among surgical abortion providers who did not see a relative advantage of medication abortion [25,43]. These factors slowed implementation of mifepristone abortion in Quebec, decreasing the VCoP's utility for physicians in that province and the likelihood that they would take the Society of Obstetricians and Gynaecologists of Canada medical abortion training program whose link to the VCoP was a primary means of recruitment [43]. Another reason for Quebec's low recruitment may relate to the province's long-standing networking organization for abortion providers, Le Comité de vigilance sur l'avortement, which meets in person 4 to 5 times a year for discussions and education on abortion and related subjects. These abortion providers may have seen few advantages to joining the CAPS-CPCA VCoP [Edith Guilbert, personal communication].

We found that knowing that there was a VCoP for support was important for many physician members when they were considering providing medication abortion. Whether this knowledge encouraged some physicians who were interested in, but uncertain about, providing medication abortion is unknown. Our qualitative data suggested that CAPS-CPCA membership increased participants' perceptions of confidence about providing abortion care, which is a determinant of adoption of new practices [53]. Aside from the reassurance of knowing that there was a place to access information and experts, our surveys and interviews as well as the website traffic and downloads indicated that members particularly valued the clinical practice tools. The large number of views and downloads of materials such as patient consent forms and information sheets, clinical checklists and guidelines, and members' comments showed the site's value as a resource repository. Although most resources were noted to be generally useful, locally relevant ones such as provincial billing codes, drug coverage, and pharmacies stocking the drug were also very important. In surveys and interviews, physician members indicated a desire for a more extensive list of pharmacies stocking the medication. Although CAPS-CPCA had more than 300 pharmacist members, fewer than 100 entered data on their pharmacy location and indicated that they had mifepristone in stock. Future VCoPs of this sort could consider approaching the large chain pharmacies for a universal input of all locations.

Logging in to the website was a deterrent for many CAPS-CPCA members, with 26.8% (34/127) of the survey respondents never doing so. To increase accessibility, we placed highly requested resources such as the clinical checklists on the landing page of the site and the very large number of views and downloads reflects the success of this strategy. Although some members found emails to be too frequent, many appreciated the emails that engaged them directly and felt that the emails built a sense of community among individuals interested in abortion. As has been shown in other research, we hypothesize that associating with other like-minded individuals may have overcome isolation and stigma that could deter some from providing abortions [29-31]. Participation in polls allowed active engagement in the VCoP to contribute data that influenced policy changes that affected members' practice, and emails kept them apprised of these changes. Although we anticipated that member involvement would diminish over time, 59.1% (75/127) of the CAPS-CPCA members who completed the 12-month survey planned to remain in the VCoP and only 10.2% (13/127) stated that they would not continue to participate.

CAPS-CPCA provided rapid access to experts for clinical questions over the study period, with 52 *Ask an Expert* and *Discussion Room* threads related to health system, logistical, and clinical support needs. Although mifepristone abortion care is usually straightforward, we were surprised that these resources were used so infrequently. The qualitative data suggested that although this feature might be important for a few providers, most would rather use or develop their own local network for expert clinical backup. Our results from interview participants who acted as mentors suggest that over time this may have occurred, with CAPS-CPCA used as a resource to support their clinical mentorship. Nevertheless, there was a desire to have clinical support or mentorship available and this was particularly valuable for new or inexperienced providers. Although challenging to achieve, linking remote or isolated clinicians with an expert mentor in their region could provide valuable, more sustainable clinical support.

Comparison With Prior Work

Similar to the findings of Carpenter et al [2] in their evaluation of learning communities in the United States, important elements of the CAPS-CPCA VCoP included credibility and trustworthiness achieved through affiliation with members' professional organizations, dissociation from the pharmaceutical industry, and leadership by known experts; active and personalized outreach to engage interested clinicians; features designed to overcome known implementation barriers and share facilitators; and responsiveness to the needs of the VCoP. We responded to early requests from busy new providers to create the clinical checklists and guidelines that became CAPS-CPCA's most viewed and downloaded resources. A qualitative study of decentralization of medication abortion services in rural Australia identified that sharing protocols and clinical resources, as we did with CAPS-CPCA, was an important enabler of clinician uptake of mifepristone abortion practice [54]. Ease of use, accessibility, and perceived usefulness have been found to be important to the success of VCoPs [2,5,55], and this was reflected in our finding that although emails and tools for practice were valued, logging in to a website was burdensome and inhibited participation for some members.

Future Directions

Sustainability of the VCoP is uncertain; continuing usefulness for members is likely to diminish as they develop experience with mifepristone abortion and connection to local experienced mentors and experts. However, with continued diffusion of this innovation, it may continue to have relevance for new abortion providers, including nurse practitioners and, potentially,

XSL•FO
RenderX

midwives. Sustainment of adoption of the innovation is an important outcome that we could not measure during our study period and an area for future research. In our interviews, lack of demand had prevented some interested physicians from implementing this practice and we hypothesize that it could similarly affect sustainment.

Our VCoP model may have applications to other clinical innovations, particularly those in focused areas where there is limited clinical experience, rapidly changing practice, unusual regulation, or associated stigma. Notably, there are established VCoPs in some jurisdictions supporting knowledge, practice, and shared experience for clinical areas such as medical assistance in dying, treatment of opioid or alcohol use disorders, and more recently COVID-19 [56-58]. Similar to CAPS-CPCA, membership in these VCoPs is not driven top-down by an organization but by individual members' interest and motivation to deliver care in these areas. For stigmatized areas of practice such as medical assistance in dying, restricting VCoP membership to ensure that members feel safe may be important. Our VCoP also kept member identities anonymous but this may have restrained social interaction. Some CAPS-CPCA members identified this as a limitation and desired local networks for personalized interaction where clinical and service issues could be discussed. Our ability to collect real-time data from our members to inform policy makers about regulatory and policy barriers to implementation was an unusual and valuable feature that could be adapted to guide health policy changes for practice improvements in other clinical areas.

Strengths and Limitations

Our research includes limitations. We were not able to isolate physician use of the website or *Member Announcements* and, thus, website data reflect use by all clinician VCoP members and Contraception and Abortion Research Team-Groupe de recherche sur l'avortement et la contraception staff who maintained the site. We believe that staff visits were most frequent in the early months when the site became active and may have artificially elevated page visits during this period. Survey and interview data provided the richest information about the physicians who joined CAPS-CPCA and its function for them as abortion providers. However, only 51.6% (222/430) and 29.5% (127/430) of the CAPS-CPCA physician members participated in the baseline and follow-up surveys, respectively. A smaller subset was invited for an interview. Physicians who were more involved in the surveys and interviews may not reflect the whole membership. To address these concerns, we purposefully invited interviewees to represent diverse perspectives, including physicians who did not join CAPS-CPCA. The strengths of this case study include the gathering of a large qualitative data set from physicians located in all areas of the country and the alignment of the findings from the website, surveys, and interviews.

Conclusions

A VCoP created to address barriers and facilitators to mifepristone abortion uptake engaged physicians from across Canada and supported some to implement this innovation in abortion practice, including those who had no previous abortion experience. Creating and widely disseminating awareness of an internet-based resource that includes practical tools for implementation, timely policy and practice updates, expert advice, and social connection may be particularly beneficial for remote and isolated providers and could encourage broader dissemination of clinical innovations.

Acknowledgments

This work was funded by the Canadian Institutes of Health Research (PHE148161 and CPP-329455-107837), the Michael Smith Foundation for Health Research (16743, 16603, and 2012-5139), and the Society of Family Planning (SFPRF11-19). SD received support as a Clinician Investigator in the Department of Family and Community Medicine, University of Toronto, and as a Scientist at the Women's College Research Institute. SM was supported as a Trainee and a Scholar of the Michael Smith Foundation for Health Research (16603 and 18270). WVN was supported as a Scholar of the Michael Smith Foundation for Health Research (2012-5139 and as an Applied Public Health Research Chair by the Canadian Institutes of Health Research (CPP-329455107837). In-kind support was contributed by the Society of Obstetricians and Gynaecologists of Canada, the College of Family Physicians of Canada, the Canadian Pharmacists Association, and the Women's Health Institute of British Columbia. The Society of Obstetricians and Gynaecologists of Canada supported development and design of the virtual community of practice platform that contributed to participant recruitment and retention through a contract with WVN's institution. All grants underwent external peer review for scientific quality, and the funders played no role in conducting the research or writing the paper.

Authors' Contributions

SD, WVN, EG, ES conceived of, and designed, the study with input from JS and SM. SD, WVN, EG, ES, CD, JS, and DJ contributed to Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement development, maintenance, and website data acquisition. SD, WVN, EG, and JS contributed to survey development. SM led the collection, analysis, and interpretation of the qualitative interviews with EG and input from SD. CD and DJ contributed to the quantitative data analysis, and SD, CD, EG, WVN, SM, DJ, and JS contributed to the interpretation. SD wrote the first draft of the manuscript, and all authors contributed to manuscript revisions and approved the final manuscript.

Conflicts of Interest

During the conduct of this study, WVN was a member of the Board of Directors of the Society of Family Planning.

XSL•FO
RenderX

Multimedia Appendix 1
Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement website screenshots.
[PPTX File , 3949 KB - jmir_v24i5e34302_app1.pptx]

Multimedia Appendix 2
Baseline questionnaire.
[DOCX File , 21 KB - jmir_v24i5e34302_app2.docx]

Multimedia Appendix 3
Follow-up questionnaire administered after 12 months.
[PDF File (Adobe PDF File), 122 KB - jmir_v24i5e34302_app3.pdf]

References

1. Ranmuthugala G, Plumb JJ, Cunningham FC, Georgiou A, Westbrook JI, Braithwaite J. How and why are communities of practice established in the healthcare sector? A systematic review of the literature. BMC Health Serv Res 2011 Oct 14;11:273 [FREE Full text] [doi: 10.1186/1472-6963-11-273] [Medline: 21999305]
2. Carpenter D, Hassell S, Mardon R, Fair S, Johnson M, Siegel S, et al. Using learning communities to support adoption of health care innovations. Jt Comm J Qual Patient Saf 2018 Oct;44(10):566-573. [doi: 10.1016/j.jcjq.2018.03.010] [Medline: 30064957]
3. Wenger E, McDermott R, Snyder WM. Cultivating Communities of Practice: A Guide to Managing Knowledge. Boston, MA, USA: Harvard Business School Press; 2002.
4. Li LC, Grimshaw JM, Nielsen C, Judd M, Coyte PC, Graham ID. Use of communities of practice in business and health care sectors: a systematic review. Implement Sci 2009 May 17;4:27 [FREE Full text] [doi: 10.1186/1748-5908-4-27] [Medline: 19445723]
5. Barnett S, Jones SC, Caton T, Iverson D, Bennett S, Robinson L. Implementing a virtual community of practice for family physician training: a mixed-methods case study. J Med Internet Res 2014 Mar 12;16(3):e83 [FREE Full text] [doi: 10.2196/jmir.3083] [Medline: 24622292]
6. Abos Mendizabal G, Nuño Solinís R, Zaballa González I. HOBE+, a case study: a virtual community of practice to support innovation in primary care in Basque Public Health Service. BMC Fam Pract 2013 Nov 05;14:168 [FREE Full text] [doi: 10.1186/1471-2296-14-168] [Medline: 24188617]
7. Sims JM. Communities of practice: telemedicine and online medical communities. Technol Forecast Soc Change 2018 Jan;126(C):53-63 [FREE Full text] [doi: 10.1016/j.techfore.2016.08.030]
8. Dressler J, Maughn N, Soon JA, Norman WV. The perspective of rural physicians providing abortion in Canada: qualitative findings of the BC Abortion Providers Survey (BCAPS). PLoS One 2013 Jun 28;8(6):e67070 [FREE Full text] [doi: 10.1371/journal.pone.0067070] [Medline: 23840588]
9. Freedman LR. Willing and Unable: Doctors' Constraints in Abortion Care. Nashville, TN, USA: Vanderbilt University Press; 2010.
10. World Health Organization. Medical management of abortion. Geneva, Switzerland: World Health Organization; 2018.
11. Mifepristone approvals. Gynuity Health Projects. 2021. URL: https://gynuity.org/assets/resources/biblio_ref_lst_mife_en.pdf [accessed 2021-07-17]
12. Contraception and Abortion Research Team-Groupe de recherche sur l'avortement et la contraception. The University of British Columbia. URL: https://cart-grac.ubc.ca [accessed 2021-06-12]
13. Norman WV, Munro S, Brooks M, Devane C, Guilbert E, Renner R, et al. Could implementation of mifepristone address Canada's urban-rural abortion access disparity: a mixed-methods implementation study protocol. BMJ Open 2019 Apr 20;9(4):e028443 [FREE Full text] [doi: 10.1136/bmjopen-2018-028443] [Medline: 31005943]
14. Canadian abortion providers support-communauté de pratique canadienne sur l'avortement (caps-cpca) community of practice. CART-GRAC (Contraception & Abortion Research Team-Groupe de recherche sur l'avortemont et la contraception). 2017. URL: https://caps-cpca.ubc.ca/index.php?title=Main_Page [accessed 2021-06-12]
15. Shaw D, Norman WV. When there are no abortion laws: a case study of Canada. Best Pract Res Clin Obstet Gynaecol 2020 Jan;62:49-62. [doi: 10.1016/j.bpobgyn.2019.05.010] [Medline: 31281015]
16. Norman WV, Downie J. Abortion care in Canada is decided between a woman and her doctor, without recourse to criminal law. BMJ 2017 Mar 24;356:j1506. [doi: 10.1136/bmj.j1506] [Medline: 28341622]
17. Mifeprex (mifepristone) information. U.S. Food & Drug Administration. 2021 Apr 31. URL: https://www.fda.gov/drugs/postmarket-drug-safety-information-patients-and-providers/mifeprex-mifepristone-information.Updated [accessed 2021-07-13]
18. Gissler M, Fronteira I, Jahn A, Karro H, Moreau C, Oliveira da Silva M, REPROSTAT group. Terminations of pregnancy in the European Union. BJOG 2012 Feb;119(3):324-332. [doi: 10.1111/j.1471-0528.2011.03189.x] [Medline: 22129480]
19. Baird B. Medical abortion in Australia: a short history. Reprod Health Matters 2015 Nov;23(46):169-176. [doi: 10.1016/j.rhm.2015.10.002] [Medline: 26719008]

XSL•FO
RenderX

20. Common drug review: clinical review report: Mifepristone and misoprostol (Mifegymiso). Canadian Agency for Drugs and Technologies in Health (CADTH). 2017 May. URL: https://www.cadth.ca/sites/default/files/cdr/pharmacoeconomic/SR0502_Mifegymiso_PE_Report.pdf [accessed 2021-09-14]
21. Munro S, Guilbert E, Wagner MS, Wilcox ES, Devane C, Dunn S, et al. Perspectives among Canadian physicians on factors influencing implementation of mifepristone medical abortion: a national qualitative study. Ann Fam Med 2020 Sep;18(5):413-421 [FREE Full text] [doi: 10.1370/afm.2562] [Medline: 32928757]
22. Endler M, Cleeve A, Sääv I, Gemzell-Danielsson K. How task-sharing in abortion care became the norm in Sweden: a case study of historic and current determinants and events. Int J Gynaecol Obstet 2020 Jul;150 Suppl 1:34-42 [FREE Full text] [doi: 10.1002/ijgo.13003] [Medline: 33219992]
23. The Abortion Act. UK Government, Public General Acts. 1967. URL: https://www.legislation.gov.uk/ukpga/1967/87/contents[Accessed [accessed 2021-02-14]
24. Mifeprex REMS Study Group, Raymond EG, Blanchard K, Blumenthal PD, Cleland K, Foster AM, et al. Sixteen years of overregulation: time to unburden Mifeprex. N Engl J Med 2017 Feb 23;376(8):790-794. [doi: 10.1056/NEJMsb1612526] [Medline: 28225670]
25. Wagner MS, Munro S, Wilcox ES, Devane C, Norman WV, Dunn S, et al. Barriers and facilitators to the implementation of first trimester medical abortion with mifepristone in the province of Québec: a qualitative investigation. J Obstet Gynaecol Can 2020 May;42(5):576-582. [doi: 10.1016/j.jogc.2019.10.037] [Medline: 31924442]
26. Health Canada updates prescribing and dispensing information for Mifegymiso. Health Canada, Government of Canada. 2017 Nov 7. URL: https://hpr-rps.hres.ca/reg-content/regulatory-decision-summary-detail.php?lang=en&linkID=RDS00294 [accessed 2021-09-30]
27. Recalls and safety alerts. Health Canada approves updates to Mifegymiso prescribing information: Ultrasound no longer mandatory. Government of Canada. 2019. URL: https://healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2019/69620a-eng.php [accessed 2021-09-20]
28. Norman WV, Guilbert ER, Okpaleke C, Hayden AS, Steven Lichtenberg E, Paul M, et al. Abortion health services in Canada: results of a 2012 national survey. Can Fam Physician 2016 Apr;62(4):e209-e217 [FREE Full text] [Medline: 28192276]
29. Freedman L, Landy U, Darney P, Steinauer J. Obstacles to the integration of abortion into obstetrics and gynecology practice. Perspect Sex Reprod Health 2010 Sep;42(3):146-151. [doi: 10.1363/4214610] [Medline: 20887283]
30. O'Donnell J, Weitz TA, Freedman LR. Resistance and vulnerability to stigmatization in abortion work. Soc Sci Med 2011 Nov;73(9):1357-1364. [doi: 10.1016/j.socscimed.2011.08.019] [Medline: 21940082]
31. Harris LH, Martin L, Debbink M, Hassinger J. Physicians, abortion provision and the legitimacy paradox. Contraception 2013 Jan;87(1):11-16. [doi: 10.1016/j.contraception.2012.08.031] [Medline: 23063339]
32. Norris A, Bessett D, Steinberg JR, Kavanaugh ML, De Zordo S, Becker D. Abortion stigma: a reconceptualization of constituents, causes, and consequences. Womens Health Issues 2011;21(3 Suppl):S49-S54. [doi: 10.1016/j.whi.2011.02.010] [Medline: 21530840]
33. Dawson AJ, Nicolls R, Bateson D, Doab A, Estoesta J, Brassil A, et al. Medical termination of pregnancy in general practice in Australia: a descriptive-interpretive qualitative study. Reprod Health 2017 Mar 14;14(1):39 [FREE Full text] [doi: 10.1186/s12978-017-0303-8] [Medline: 28288649]
34. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
35. Rogers EM. Diffusion of Innovation. 5th edition. New York, NY, USA: Simon & Schuster; 2003.
36. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
37. Munro S, Wahl K, Soon JA, Guilbert E, Wilcox ES, Leduc-Robert G, et al. Pharmacist dispensing of the abortion pill in Canada: diffusion of innovation meets integrated knowledge translation. Implement Sci 2021 Aug 03;16(1):76 [FREE Full text] [doi: 10.1186/s13012-021-01144-w] [Medline: 34344393]
38. Bowen S, Graham ID. Integrated knowledge translation. In: Strauss SE, Tetroe J, Graham ID, editors. Knowledge Translation in Health Care. Oxford, UK: Wiley Online Library; 2013:14-23.
39. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2022-04-22]
40. Rebic N, Munro S, Norman WV, Soon JA. Pharmacist checklist and resource guide for mifepristone medical abortion: user-centred development and testing. Can Pharm J (Ott) 2021 May 28;154(3):166-174 [FREE Full text] [doi: 10.1177/17151635211005503] [Medline: 34104270]
41. Medical Abortion Training Program. The Society of Obstetricians and Gynaecologists of Canada. URL: https://www.sogc.org/en/rise/Events/event-display.aspx?EventKey=MATP [accessed 2021-10-17]
42. Devane C, Renner RM, Munro S, Guilbert É, Dunn S, Wagner MS, et al. Implementation of mifepristone medical abortion in Canada: pilot and feasibility testing of a survey to assess facilitators and barriers. Pilot Feasibility Stud 2019 Nov 8;5:126 [FREE Full text] [doi: 10.1186/s40814-019-0520-8] [Medline: 31720004]

XSL•FO
RenderX

43. Guilbert E, Wagner MS, Munro S, Wilcox ES, Dunn S, Soon JA, et al. Slow implementation of mifepristone medical termination of pregnancy in Quebec, Canada: a qualitative investigation. Eur J Contracept Reprod Health Care 2020 Jun;25(3):190-198. [doi: 10.1080/13625187.2020.1743825] [Medline: 32312130]
44. What is AWStats. AWStats. URL: https://awstats.sourceforge.io/ [accessed 2022-01-19]
45. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
46. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
47. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597 [FREE Full text] [doi: 10.1080/2159676x.2019.1628806]
48. Hesse-Biber S. Feminist approaches to triangulation: uncovering subjugated knowledge and fostering social change in mixed methods research. J Mix Methods Res 2012 Mar 30;6(2):137-146. [doi: 10.1177/1558689812437184]
49. Population estimates, July 1, by census metropolitan area and census agglomeration, 2016 boundaries. Statistics Canada. 2021 Sep 12. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710013501 [accessed 2021-07-23]
50. Annual Demographic Estimates: Canada, Provinces and Territories, 2018 (Total Population only). Statistics Canada. 2018 Sep 27. URL: https://www150.statcan.gc.ca/n1/pub/91-215-x/2018001/sec1-eng.htm [accessed 2021-08-23]
51. Collège des médecins du Québec. URL: http://www.cmq.org/publications-pdf/p-6-2015-01-07-fr-code-de-deontologie-des-medecins.pdf?t=1645131398609 [accessed 2022-02-17]
52. Norman WV and members of CART-GRAC. Understanding Medical Abortion Access and Safety in Canada: Findings of the Health Canada collaboration with a researcher-network for post-marketing surveillance of Mifegymiso. Ottawa, ON, Canada: Health Canada's Chief Medical Officer and her division head reports, Government of Canada; Feb 25, 2021. URL: https://med-fom-cart-grac.sites.olt.ubc.ca/files/2022/04/ExecSummary_CART-MifeStudies_2021-02-24.pdf
53. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. Implement Sci 2008 Jul 16;3:36 [FREE Full text] [doi: 10.1186/1748-5908-3-36] [Medline: 18631386]
54. Hulme-Chambers A, Clune S, Tomnay J. Medical termination of pregnancy service delivery in the context of decentralization: social and structural influences. Int J Equity Health 2018 Nov 21;17(1):172 [FREE Full text] [doi: 10.1186/s12939-018-0888-8] [Medline: 30463561]
55. Jiménez-Zarco AI, González-González I, Saigí-Rubió F, Torrent-Sellens J. The co-learning process in healthcare professionals: assessing user satisfaction in virtual communities of practice. Comput Hum Behav 2015 Oct;51:1303-1313 [FREE Full text] [doi: 10.1016/j.chb.2014.11.057]
56. Medical assistance in dying. Canadian Association of MAID Assessors and Providers (CAMAP/ACEPA). URL: https://camapcanada.ca/# [accessed 2021-07-21]
57. Atlantic Mentorship Network - Pain & Addiction. URL: https://www.atlanticmentorship.com/ [accessed 2021-07-21]
58. COVID-19 community of practice for Ontario family physicians. Ontario College of Family Physicians and University of Toronto Department of Family and Community Medicine. URL: https://www.ontariofamilyphysicians.ca/tools-resources/covid-19-resources/covid-19-community-of-practice-for-ontario-family-physicians [accessed 2021-07-21]

Abbreviations

CAPS-CPCA: Canadian Abortion Providers Support-Communauté de pratique canadienne sur l'avortement
CoP: community of practice
VCoP: virtual community of practice

Edited by T Leung; submitted 18.10.21; peer-reviewed by H Oh, C Urquhart, JW Dearing, A Hart; comments to author 25.12.21; revised version received 17.02.22; accepted 24.03.22; published 05.05.22.

Please cite as:
Dunn S, Munro S, Devane C, Guilbert E, Jeong D, Stroulia E, Soon JA, Norman WV
A Virtual Community of Practice to Support Physician Uptake of a Novel Abortion Practice: Mixed Methods Case Study
J Med Internet Res 2022;24(5):e34302
URL: https://www.jmir.org/2022/5/e34302
doi:10.2196/34302
PMID:35511226

©Sheila Dunn, Sarah Munro, Courtney Devane, Edith Guilbert, Dahn Jeong, Eleni Stroulia, Judith A Soon, Wendy V Norman. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 05.05.2022. This is an open-access

XSL•FO
RenderX

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

XSL•FO
RenderX

Original Paper

Research on Digital Technology Use in Cardiology: Bibliometric Analysis

Andy Wai Kan Yeung[1,2], BDS, PhD; Stefan Tino Kulnik[3], PhD; Emil D Parvanov[2,4], PhD; Anna Fassl[2], MA; Fabian Eibensteiner[2,5], MD; Sabine Völkl-Kernstock[2], PhD; Maria Kletecka-Pulker[2,6], PhD; Rik Crutzen[3,7], PhD; Johanna Gutenberg[3,7], MSc; Isabel Höppchen[3,8], MSc; Josef Niebauer[3,9,10], MBA, MD, PhD; Jan David Smeddinck[3], PhD; Harald Willschke[2,11], MD; Atanas G Atanasov[2,12], PhD

[1]Division of Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

[2]Ludwig Boltzmann Institute Digital Health and Patient Safety, Medical University of Vienna, Vienna, Austria

[3]Ludwig Boltzmann Institute for Digital Health and Prevention, Salzburg, Austria

[4]Department of Translational Stem Cell Biology, Research Institute of the Medical University of Varna, Varna, Bulgaria

[5]Division of Pediatric Nephrology and Gastroenterology, Department of Pediatrics and Adolescent Medicine, Comprehensive Center for Pediatrics, Medical University of Vienna, Vienna, Austria

[6]Institute for Ethics and Law in Medicine, University of Vienna, Vienna, Austria

[7]Department of Health Promotion, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

[8]Center for Human Computer Interaction, Paris Lodron University Salzburg, Salzburg, Austria

[9]University Institute of Sports Medicine, Prevention and Rehabilitation, Paracelsus Medical University Salzburg, Salzburg, Austria

[10]REHA Zentrum Salzburg, Salzburg, Austria

[11]Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University Vienna, Vienna, Austria

[12]Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Jastrzebiec, Poland

Corresponding Author:
Atanas G Atanasov, PhD
Ludwig Boltzmann Institute Digital Health and Patient Safety
Medical University of Vienna
Spitalgasse 23, Bauteil 86, 2 Stock
Vienna, 1090
Austria
Phone: 43 664 1929 85
Email: atanas.atanasov@univie.ac.at

Abstract

Background: Digital technology uses in cardiology have become a popular research focus in recent years. However, there has been no published bibliometric report that analyzed the corresponding academic literature in order to derive key publishing trends and characteristics of this scientific area.

Objective: We used a bibliometric approach to identify and analyze the academic literature on digital technology uses in cardiology, and to unveil popular research topics, key authors, institutions, countries, and journals. We further captured the cardiovascular conditions and diagnostic tools most commonly investigated within this field.

Methods: The Web of Science electronic database was queried to identify relevant papers on digital technology uses in cardiology. Publication and citation data were acquired directly from the database. Complete bibliographic data were exported to VOSviewer, a dedicated bibliometric software package, and related to the semantic content of titles, abstracts, and keywords. A term map was constructed for findings visualization.

Results: The analysis was based on data from 12,529 papers. Of the top 5 most productive institutions, 4 were based in the United States. The United States was the most productive country (4224/12,529, 33.7%), followed by United Kingdom (1136/12,529, 9.1%), Germany (1067/12,529, 8.5%), China (682/12,529, 5.4%), and Italy (622/12,529, 5.0%). Cardiovascular diseases that had been frequently investigated included hypertension (152/12,529, 1.2%), atrial fibrillation (122/12,529, 1.0%), atherosclerosis (116/12,529, 0.9%), heart failure (106/12,529, 0.8%), and arterial stiffness (80/12,529, 0.6%). Recurring modalities were electrocardiography (170/12,529, 1.4%), angiography (127/12,529, 1.0%), echocardiography (127/12,529, 1.0%), digital subtraction

XSL•FO
RenderX

angiography (111/12,529, 0.9%), and photoplethysmography (80/12,529, 0.6%). For a literature subset on smartphone apps and wearable devices, the Journal of Medical Internet Research (20/632, 3.2%) and other JMIR portfolio journals (51/632, 8.0%) were the major publishing venues.

Conclusions: Digital technology uses in cardiology target physicians, patients, and the general public. Their functions range from assisting diagnosis, recording cardiovascular parameters, and patient education, to teaching laypersons about cardiopulmonary resuscitation. This field already has had a great impact in health care, and we anticipate continued growth.

(J Med Internet Res 2022;24(5):e36086) doi:10.2196/36086

KEYWORDS

cardiovascular; heart; hypertension; atrial fibrillation; cardiopulmonary resuscitation; electrocardiography; photoplethysmography; wearable device, digital health, mHealth; cardiology; cardiac; health application

Introduction

Background

Modern health care and medicine are characterized by continuous digital innovation. This innovation is driven by the confluence of, first, technological advances with transformative potential and, second, convincing use cases based on needs and opportunities from the health care domain. This is an area of high-volume activity evidenced in a large and heterogeneous scientific literature base, which warrants a high-level overview and bibliometric analysis.

Current Transformative Developments in Digital Technology

Recent advances in digital technology for health care and medicine have been fundamentally facilitated by a revolution—increasing miniaturization and affordability—in sensing devices, which have been manufactured as both stationary and wearable devices to track a broad and growing range of vital signs and physiological measurements [1,2]. These developments have coincided with rapid innovations in interactive, networked, mobile, and ubiquitous computing [3], which has brought about modern smartphones, wireless connectivity, and Internet of Things computing, networked information systems, and increasingly capable consumer--facing and professional apps [4]. This enables effective automation in many areas that are highly relevant for health care and medicine, such as communication (eg, telehealth [5], which has been recently emphasized by an increased need for remote access for medical support in both physical and mental health during the COVID-19 pandemic [6,7]), social support [8-10], and education [11]. Moreover, there are growing possibilities for the augmentation of sensing and actuation [12], via biocompatible technologies [13] and ubiquitous sensing focused on situated functionality [14].

Technology transfer in these areas follows a general pathway from innovators and early adopters—technology developments are often inspired by hacking, gaming, or similar communities—through applied research and development into actual medical and health care practice [15]. Virtual, augmented, and mixed reality are good examples of current technologies that are beginning to take hold in real-life medical and health care practice, for example, in diagnostic and surgical procedures and rehabilitation, by offering versatility for a broad range of conditions, including pain, stroke, anxiety, depression, fear, cancer, and neurodegenerative disorders [16].

Other recent developments with transformative potential include initiatives toward digital biomarkers and interventions that promise to enable personalized and precision medicine [17]. Building on foundations developed in enthusiast communities around the quantified self [18] and personal informatics [19-21], these approaches suggest there is a need for patient data contributions and personal health records [22,23] with advances in data processing and analytics, for example, in artificial intelligence and machine learning for supporting diagnosis [24,25] and medical decision-making [26,27]. Key drivers toward truly personalized and precision medicine [28,29] will arguably be the adaptability and adaptivity of systems that anticipate rather than react [30,31], for example, via predictive modeling [32], which in turn facilitates a focus toward preventative rather than curative medicine [33].

Further potentially transformative technologies are conversational interfaces [34-36], developments that enable localized and individualized production through 3D printing [37,38], biochemical composition [39], or personal genomics [40]. These developments have considerable potential for positive change but also require delicate handling of personal data and privacy issues in accordance with data standards [41], legal and ethical considerations [42,43], and social considerations [44,45]. A key challenge lies in moving toward more sustainable adoption and use of available technologies, which requires a broad view on complex ecosystems [46,47], motivation [48] and habituation or behavior change [49-51]. Moreover, there is a need to more closely connect research and industry [52] and to work in a highly human-centered manner [53].

Clinical Use Cases of Digital Innovation in Cardiology

The variety of digital technologies in health care and medicine is reflected in the field of cardiology, in which multiple uses can be found. Telecardiology describes the delivery of one-to-one cardiology care without the need for physical meetings between the physician and the patient [54] and has been facilitated by the improved availability and functionality of remote communication technologies and by digital technologies that enable reliable recording and transmission of clinical measurements from implantable (pacemakers, defibrillators) and consumer devices (blood pressure monitors, scales, thermometers) [54]. Cardiac telerehabilitation—programs

XSL•FO
RenderX

provided at patients' homes rather than at rehabilitation centers [55]—uses technology solutions to facilitate the remote instruction, monitoring, and supervision of patients during exercise training, with processes for providing emergency care in case of medical emergencies [56]. Artificial intelligence and machine learning approaches offer multitudes of possibilities in cardiology diagnostics and therapeutics, for example, individual cardiovascular risk factor identification; profiling, prediction, and management of cardiac arrhythmias; and enhanced cardiac imaging [17,57,58].

The field of behaviour change for primary and secondary prevention of cardiovascular disease through digital technologies, for example, to understand and modify behavior, has increased rapidly in recent years [59]. This approach could deliver effective personalized support for heart-healthy lifestyle changes, such as adherence to medication and exercise recommendations [17] with the measurement of physical activity [60] and associated parameters such as heart rate [61] using sensors incorporated in objects of daily use, such as mobile phones and watches [62]. Existing technologies also provide the ability to capture information about the environment in which behavior takes place, with mobile phone location tracking, and can be used to facilitate understanding of behavior [63] or to change behavior [64]. Behavior change interventions can be effective, especially when tailored to the individual [65]. However, there is room for improvement in terms of using the unique characteristics and full potential of digital technologies, such as the possibility of intervening at the right moment (for example when a person is in need of support). The potential of these so-called just-in-time adaptive interventions has only been explored recently, and insight into their effectiveness is largely still lacking [66].

Given that digital technologies (specifically, the internet and smartphone apps) are vehicles for information transfer, another highly promising area of application for these digital technologies in cardiology is health literacy (ie, the degree to which individuals can obtain, process, understand, and communicate about health-related information needed to make informed health decisions [67]). Health literacy is a prerequisite to successfully maintain health and self-care; navigate through the healthcare system; and in case of illness, understand health information, medication, and treatment plans [68]. Especially in older adults, health literacy is a significant predictor of information-seeking behaviors and health outcomes [68]. Despite growing global recognition of health literacy as a critical determinant of health and well-being and efforts to improve health literacy [69], health literacy levels among the global population remain low [70-72].

Digital technologies, including the internet and information communication technologies, seem to offer a convenient way to deliver broadly and rapidly evidence-based health information and thus improve overall health literacy, especially in disadvantaged populations that lack access to health care and relevant health information [73,74]. However, a recent study [75] has shown that persons with lower health literacy report difficulties searching health information and are less likely to use search engines. Moreover, low health literacy is also associated with difficulty judging the quality of health information from the internet [76]. In order to actively support individuals' health literacy, digital technologies or services are increasingly promoted in different care contexts to accelerate patient–provider communication and, at the same time, offer an opportunity to educate patients in the appropriate use of web-based health information. In inpatient care, digital tools such as electronic displays can be employed during ward rounds to support the consultation or facilitate the discharge process, and medically vetted electronic health information is shared with patients at the hospital bedside [77]. Automated systems can be integrated to teach patients about their diagnosis and postdischarge self-care regimen [78]. In outpatient care, digital technologies often aim to support chronically ill persons. Telehealth systems for synchronous audio- and video-based communication allow patients to report symptoms and preferences to their health care provider remotely [79], while asynchronous text-based communication through patient portals enables patients more convenient access to their health information [80]. These technologies offer patients alternative modalities for information transfer and communication with health care providers, thereby facilitating effective information exchange and supporting individual health literacy skills. In the field of cardiology, the importance of a greater focus on supporting health literacy has recently been highlighted, specifically in the context of primary and secondary prevention of cardiovascular disease [81].

Rationale for a Bibliometric Analysis

The broad range of digital technology use in cardiology is reflected by a large scientific literature base. Bibliometric analysis provides an integral view with quantitative evaluations of publishing metrics of research literature [82-84]. The purpose of this bibliometric analysis of digital technology uses in cardiology is to describe and discover current trends, topics, and scientometric characteristics within this body of literature, providing a high-level overview of the scientific literature and enabling insights for future directions in digital health in cardiology. To the best of our knowledge, no such analysis has been published to date.

Methods

We searched the Web of Science Core Collection database on November 22, 2021 (Textbox 1).

We excluded *digitalin**, *digitalis**, *supplemental digital* and *digital ulcer** because these words and their derivatives did not refer to digital technology, but instead referred to the drug *digitalin*, to the plant genus *Digitalis*, to supplemental digital content, and to the medical condition *digital ulcer*, respectively. No additional filters were applied to restrict the search results. The search resulted in 12,529 papers. The *Analyze Results* and *Citation Report* functions of the Web of Science platform were utilized for basic frequency counts and the number of citations per publication (mean citations per item within a subset) of the most productive authors, institutions, countries, journals, and journal categories. We also defined a subset—literature that included the terms *smartphone**, *app*, or *wearable** in the title, abstract, or keywords—which contained 632 papers.

XSL•FO
RenderX

The full record and cited references were then exported into VOSviewer as tab delimited files to synthesize a term map. For clarity, only terms that appeared in at least 0.5% of the literature set (>63) were included in the map. A list of top 5000 common words from the Corpus of Contemporary American English was entered to remove generic (and therefore, less meaningful) words from the term map [85]. VOSviewer was also used to identify the top 20 recurring author keywords.

As the latest digital technology uses often involve smartphone apps and wearable devices, This analysis described above, except for the term map, was similarly conducted on a subset of the concerned.

Textbox 1. Digital technology in cardiology search string. TS: searching for title, abstract, and keywords; WC: searching for the particular journal category.

(#1 OR #2) NOT (#3 OR #4)

where

1. TS=(digital* AND (cardio* OR cardiac* OR heart*) NOT (digitalin* OR digitalis*))
2. WC=(CARDIAC CARDIOVASCULAR SYSTEMS) AND TS=(digital* NOT (digitalin* OR digitalis*))
3. TS=("Supplemental Digital")
4. TS=("digital ulcer*")

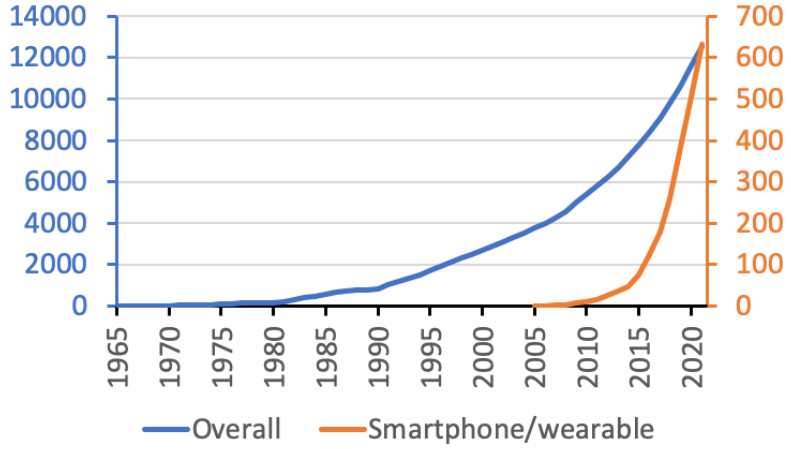
Results

The 12,529 papers were published from 1965 to November 22, 2021. The earliest publication was a report on the development and demonstration of an analog-digital analyzing unit to screen heart sounds in children [86]. The literature growth seemed to be accelerating in the 2000s and especially into the 2010s (Figure 1). Approximately three-quarters of the papers (9271/12,529, 74.0%) were original articles, and review papers accounted for 6.3% (789/12,529). Proceedings papers and meeting abstracts accounted for 14.2% (1779/12,529) and 6.0% (752/12,529), respectively.

The most productive author was Professor David J Sahn from Oregon Health and Science University, whose highly cited papers were focused on real-time 3D echocardiography [87-89]. Of the 5 most productive institutions, 4 were based in the United States of America, with Harvard University having the highest number of citations per publication. The most productive journals were from the area of cardiology or cardiovascular system, with *Circulation* having the highest citations per publication among the top 5 (Table 1). *Cardiac and cardiovascular systems* was the most productive journal category, accounting for nearly one-third of the papers.

The variety of digital technology uses in cardiology can be observed (Figure 2), with uses related to blood pressure (n=727, citations per publication: 20.1), hypertension (n=642, citations per publication: 21.1), arterial stiffness (n=128, citations per publication: 19.4), and stenosis (n=500, citations per publication: 23.7). Terms that appeared in more recent papers included *wearable device* (n=79, citations per publication: 10.1), *smartphone* (n=143, citations per publication: 12.0), and *COVID* (n=111, citations per publication: 3.2) (Figure 2), as well as *pandemic* (n=74, citations per publication: 3.0), *machine learning* (n=97, citations per publication: 11.1), *artificial intelligence* (n=112, citations per publication: 8.8), and *app* (n=149, citations per publication: 10.3) (not in Figure 2).

Figure 1. Papers published on digital technology uses in cardiology.



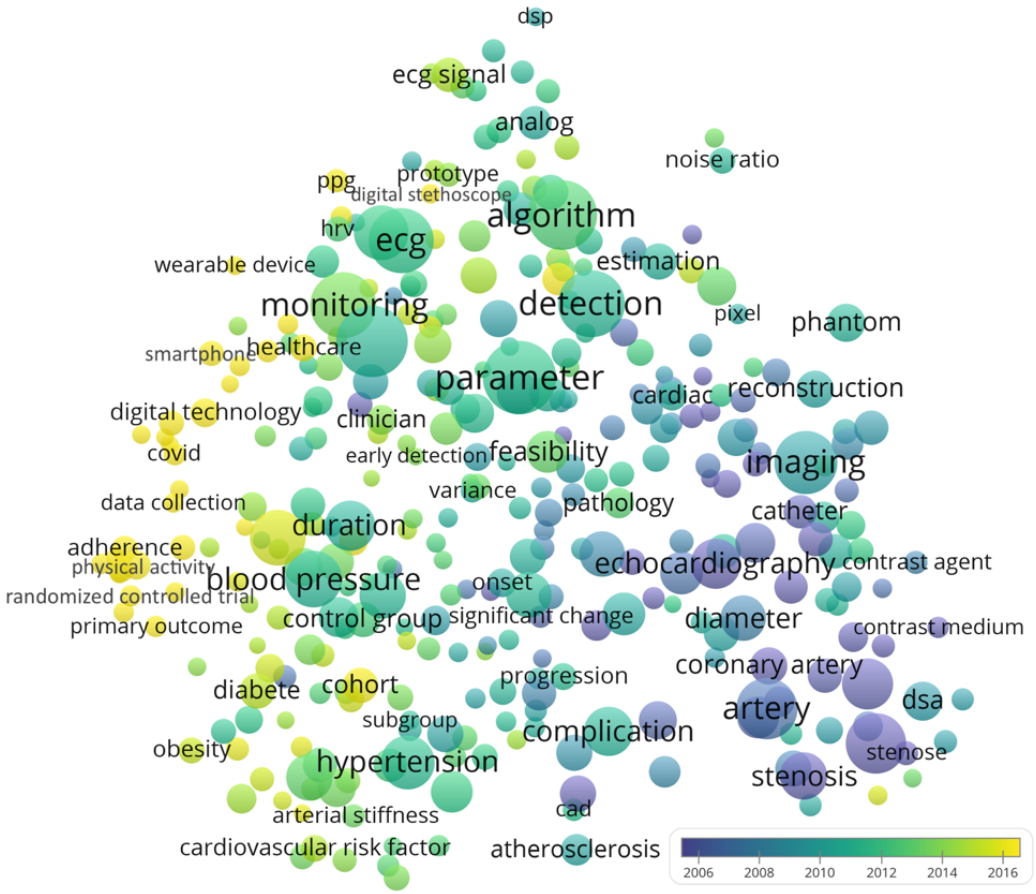
XSL•FO
RenderX

Table 1. Top 5 (most productive) entities in literature on digital technology use in cardiology.

Entity	N (%)	Citations per publication
Author		
Sahn, David J	71 (0.5)	10.7
Wong, Tien Yin	53 (0.5)	46.2
Molloi, Sabee	35 (0.3)	14.9
Jones, Molly	33 (0.3)	14.0
Li, Xiang-Ning	33 (0.3)	4.5
Institution		
University of California system	440 (3.5)	32.2
University of London	278 (2.2)	30.8
Harvard University	263 (2.1)	54.4
Duke University	178 (1.4)	35.4
US Department of Veterans Affairs	163 (1.3)	44.0
Country		
United States of America	4224 (33.7)	26.7
United Kingdom	1136 (9.1)	24.8
Germany	1067 (8.5)	23.7
China	682 (5.4)	12.2
Italy	622 (5.0)	18.4
Journal		
Circulation	411 (3.3)	57.4
Journal of the American College of Cardiology	271 (2.2)	34.4
Cardiovascular and Interventional Radiology	251 (2.0)	13.8
American Journal of Cardiology	171 (1.4)	29.3
European Heart Journal	140 (1.1)	20.2
Journal category		
Cardiac cardiovascular systems	4101 (32.7)	22.0
Radiology, nuclear medicine, or medical imaging	1289 (10.3)	20.9
Peripheral vascular disease	1090 (8.7)	36.3
Engineering, biomedical	1001 (8.0)	20.5
Engineering, electrical or electronic	965 (7.7)	9.3

XSL•FO
RenderX

Figure 2. Recurring terms in the titles and abstracts of the literature about digital technology applications in cardiology. Circle size indicates publication count. Circle color indicates the average publication year. Distances between circles indicate how frequently the terms co-occurred.



The terms *telemedicine*, *digital health*, and *mHealth (mobile health)* were among the top 20 author keywords (Table 2), suggesting that digital technology uses have been a major research focus. Such uses are also of increasing interest in the context of the COVID-19 pandemic.

The literature subset contained 632 papers on smartphone apps and wearable devices. The first paper of this subset was published in 2005, and it introduced a wearable multiparameter (including heart rate and blood pressure) ambulatory physiological monitoring system that could digitally record and continuously stream data to a base station [90]. Approximately 62.2% (393/632) of papers were original articles. Review papers accounted for 15.2% (96/632). The most productive author within this subset was Dr. Mohamed Elgendi affiliated with University of British Columbia and Simon Fraser University. His research interest focused on using data from photoplethysmography (PPG) to detect hypertension, potentially with the aid of machine learning [91,92]. Of the 5 most productive institutions, 4 were based in the United States of America. Furthermore, in the top 5 countries, the first 4 places remained unchanged from the those of the full data set (United States, United Kingdom, Germany, and China). Fifth place was taken by Australia, while Italy (ranked as the fifth most productive country in the full data set) moved to tenth place. JMIR Publications had the top 3 journals, which collectively accounted for more than 10% of the 632 papers (Table 3).

The top 20 author keywords (Table 4), notably, included *atrial fibrillation*, a cardiac condition that causes rapid and irregular heart rate.

XSL•FO
RenderX

Table 2. Top 20 author keywords for digital technology use in cardiology.

Author keyword	n (%)	Citations per publication
Heart rate	173 (1.4)	56.5
ECG (electrocardiography)	170 (1.4)	9.6
Telemedicine	169 (1.3)	14.8
Digital health	167 (1.3)	10.7
Blood pressure	159 (1.3)	18.4
Hypertension	152 (1.2)	17.6
Angiography	127 (1.0)	17.0
Echocardiography	127 (1.0)	38.6
Heart rate variability	126 (1.0)	18.1
Atrial fibrillation	122 (1.0)	20.6
Atherosclerosis	116 (0.9)	21.3
Cardiovascular disease	115 (0.9)	16.2
Digital subtraction angiography	111 (0.9)	12.5
Coronary artery disease	107 (0.9)	28.1
Heart failure	106 (0.8)	14.7
Machine learning	97 (0.8)	4.8
Heart	92 (0.7)	23.2
mHealth	89 (0.7)	13.7
Photoplethysmography	80 (0.6)	30.0
Arterial stiffness	80 (0.6)	22.6

XSL•FO
RenderX

Table 3. Top 5 in the literature subset (literature related to smartphone apps and wearable devices). The author list contains more than 5 names since multiple authors had the same number of papers.

Entity	n (%)	Citations per publication
Author (last name, first name)		
Elgendi, Mohamed	9 (1.4)	18.4
Martin, Seth S	8 (1.3)	22.5
Sharma, Abhinav	6 (0.9)	15.3
Ward, Rabab	6 (0.9)	23.0
Benjamin, Emelia J.	5 (0.8)	20.2
Majmudar, Maulik	5 (0.8)	4.0
Marvel, Francoise A	5 (0.8)	4.0
Murabito, Joanne M	5 (0.8)	20.2
Shan, Rongzi	5 (0.8)	10.0
Tarakji, Khaldoun G	5 (0.8)	42.4
Van Hoof, Chris	5 (0.8)	13.4
Institution		
University of California system	31 (4.9)	18.4
Harvard University	27 (4.3)	10.6
Stanford University	21 (3.3)	29.4
University of London	18 (2.8)	14.1
Johns Hopkins University	15 (2.4)	13.3
Country		
United States of America	242 (38.3)	15.8
United Kingdom	59 (9.3)	11.2
Germany	51 (8.1)	7.6
China	40 (6.3)	17.7
Australia	37 (5.9)	8.5
Journal		
JMIR mHealth and uHealth	30 (4.7)	5.8
JMIR Research Protocols	21 (3.3)	2.3
Journal of Medical Internet Research	20 (3.2)	11.5
Sensors	19 (3.0)	13.1
IEEE Access	10 (1.6)	7.8
Journal category		
Engineering Electrical Electronic	135 (21.4)	8.4
Health Care Sciences Services	104 (16.5)	8.5
Medical Informatics	90 (14.2)	10.7
Cardiac Cardiovascular Systems	89 (14.1)	12.0
Engineering Biomedical	69 (10.9)	12.6

XSL•FO
RenderX

Table 4. Top 20 author keywords of the literature subset (literature related to smartphone apps and wearable devices).

Author keyword	n (%)	Citations per publication
Digital health	72 (11.4)	14.1
mHealth (mobile health)	47 (7.4)	12.5
Wearables	41 (6.5)	9.9
Smartphone	36 (5.7)	10.5
Telemedicine	28 (4.4)	10.0
Wearable	24 (3.8)	11.0
Heart rate	22 (3.5)	7.2
Mobile phone	21 (3.3)	8.9
Wearable devices	20 (3.2)	11.6
ECG (electrocardiography)	20 (3.2)	5.9
Physical activity	18 (2.8)	13.4
Digital medicine	17 (2.7)	18.1
Machine learning	17 (2.7)	11.9
Cardiovascular disease	16 (2.5)	14.3
Stress	16 (2.5)	3.5
Heart rate variability	15 (2.4)	6.7
Atrial fibrillation	14 (2.2)	20.9
Cardiology	14 (2.2)	4.4
Artificial intelligence	14 (2.2)	4.2
eHealth	13 (2.1)	16.1

Discussion

Cardiovascular diseases that were frequently indicated as author keywords in the 12,529 papers included hypertension, atrial fibrillation, atherosclerosis, heart failure, and arterial stiffness. A recent meta-analysis reported that using smartphone app–based interventions could significantly lower blood pressure and improve medication adherence in patients with hypertension [93]. It was found that both wearable, ambulatory, and home monitoring devices recorded blood pressure with comparable values [94]. Smartphone and smartwatch apps could already readily distinguish atrial fibrillation from sinus rhythm and detect them with high sensitivity and specificity comparable to 12-lead electrocardiography (ECG) [95,96]. Authoritative bodies such as the European Society of Cardiology have also developed smartphone apps for patient education on atrial fibrillation [97]. The use of smartphone apps could help general physicians and trainee cardiologists decide whether a patient with heart failure should receive an implantable cardioverter defibrillator or cardiac resynchronization therapy [98]. Researchers found that these apps could potentially reduce hospital staff and facility costs by enabling patients to self-perform simple diagnostic tests, such as the 6-minute walk test, a functional exercise test used to assess patients with cardiopulmonary problems [99]. Similarly, improved access and participation in cardiac rehabilitation in terms of physical activity counselling and exercise training could be achieved by using digital health interventions that were not facility-based [100]. Apart from patient and physicians, digital technology could also target people outside of health care. For instance, massively multiplayer virtual worlds could be modified for use as a serious game to efficiently and reliably teach high school students how to perform cardiopulmonary resuscitation, an act that can be life-saving [101]. Virtual reality, a research hotspot in recent years [16], could also be utilized to teach cardiopulmonary resuscitation for medical students [102].

Meanwhile, recurring investigative modalities highlighted by the current analysis included ECG, angiography, echocardiography, digital subtraction angiography, and PPG. PPG is one of the most heavily researched diagnostic tools, and it is noninvasive, inexpensive, and convenient [103]. It could also be performed with a smartphone to detect heart rate with an average error rate as low as 1 to 1.5% [104]. Applying deep learning to PPG data could also stratify patients' risk of hypertension [92]. Moreover, artificial intelligence could interpret ECGs rapidly with human-like performance and even detect signals and patterns largely unrecognizable by humans [105]. Overall, use cases, in which physiological parameters from wearable sensing devices are extracted and artificial intelligence is applied to draw insights, are a focal point in the literature; there is a large cluster of prominent terms such as *parameter*, *monitoring*, *detection*, and *algorithm* (Figure 2). Machine learning methods such as deep learning are frequently used to represent data structures and to make predictions or classifications, with the overall intention of supporting clinicians in data-based decision-making [106]. The expectation is that

XSL•FO
RenderX

this will contribute to increasing the efficiency and effectiveness of care delivery, in particular with respect to precision health and personalized care [17]. In fact, digital technology could be very useful, with predictive models and interventions in the personalized management of cardiovascular disease patients for predicting sudden cardiac death, ventricular tachycardia, and ventricular fibrillation [107,108].

During the COVID-19 pandemic, the value of digital technology use under extreme measures for infection control has become evident. For instance, electronic stethoscopes could be utilized for contactless auscultation with real-time playback, digital storage of data, and subsequent data transmission for further assessment [109]. With the reduction of in-person hospital visits, digital technology could facilitate telemonitoring programs to serve as alternative to support patient access to care [110]. Indeed, a recent bibliometric analysis on digital health papers listed *telemedicine* and *telehealth* as two of the most frequently used keywords, indicating their relevance beyond cardiology [111].

In the subset of smartphone app and wearable device literature, we found that *mHealth*, *physical activity*, and *eHealth* were among the top author keywords, and most papers had been published in *JMIR mHealth and uHealth*, *JMIR Research Protocols*, and *Journal of Medical Internet Research*. These findings were highly consistent with a recent bibliometric analysis on digital health behavior change technology [59], but where United States, United Kingdom, and the Netherlands had been the most productive countries, in our findings, the Netherlands was replaced by Germany and China. This suggests that there are some geographical differences in research interest between cardiology-specific and general research on health behavior change. Meanwhile, another recent bibliometric analysis on mobile health apps also identified the 3 abovementioned journals as the most productive [112].

In principle, smartphone apps could offer an ideal modality for delivering digital interventions to empower patients' self-management, by providing health literacy support and coaching content (eg, a smartphone coaching app for blood pressure control [113]). However, in line with findings from this bibliometric analysis, recent reviews [114,115] have highlighted that there is a relative paucity of health literacy interventions and, more specifically, a paucity of digital health literacy interventions for cardiovascular patient groups [116]. Moreover, apps designed to empower patients often include a narrow range of features and lack explicit linkage with theories of empowerment [117,118]. This is an area for further research—the development of content and features for such apps should be based on relevant theoretical underpinnings.

Interestingly, this bibliometric analysis did not identify top-listed terms related to *primary prevention* or *secondary prevention/cardiac rehabilitation* of cardiovascular disease. This may seem surprising, since there has been a rapid growth in the development of health apps and other digital technology interventions, and primary and secondary prevention of cardiovascular disease and cardiac rehabilitation are important areas of application [119-121]. A number of recent reviews demonstrate that a sizeable body of literature is available, for example, a systematic review and meta-analysis [122], which included 51 primary studies of digital health interventions for the primary and secondary prevention of cardiovascular disease; a systematic review and meta-analysis [123], which reported on 25 original studies of digital technology interventions for cardiovascular risk factor modification; a scoping review [124], which summarized 13 trials of mobile technology interventions for improving exercise capacity in cardiac rehabilitation; and a systematic review [100], which reported on 31 primary studies of digital health interventions for physical activity and exercise adherence in cardiac rehabilitation. In the context of this bibliometric analysis, this indicates that the literature on digital technology cardiology uses appears more accessible through disease- and condition-specific key terms (hypertension, atrial fibrillation, atherosclerosis, cardiovascular disease, and heart failure), as opposed to more service- and patient pathway–oriented terms (primary prevention, secondary prevention, and cardiac rehabilitation), which may be a relevant consideration in designing literature search strategies for researchers targeting the latter [107,108].

Another aspect of digital technology use in cardiology that was not featured prominently among the findings of this bibliometric analysis is the use of digital technology to increase the efficiency and quality of research in cardiology [125]. This refers to new possibilities afforded by mobile apps, smart devices, and implantable or wearable technologies for the design and management of research studies. Digital processes for data collection, monitoring, communication, documentation and approvals in research hold potential cost and time savings, and functionalities of digital devices open new avenues in the collection and quality control of real-time continuous data acquisition [125] (eg, real-time capture of self-reported measures and symptoms in web-based forms, and the verification of subjective data through concurrent objective measurement, for example, by supplementing subjective reports of physical activity with continuously worn activity tracking devices). The use of digital technologies in the design and management of research studies in cardiology is an emerging focus in the literature, with opportunities for robust evaluations of the advantages of digital research designs over traditional nondigital approaches.

We observed that cardiology journals were predominant in the entire literature set. The top 5 journals were also among those that had published the 100 most cited cardiovascular papers in a previous study [126], with *Circulation* and *European Heart Journal* together accounting for 64% of the top 100. However, it should be noted that digital technology use does not only involve cardiology but is an intersection between medical informatics, engineering and health sciences and services in general. With this in mind, when the literature subset on smartphone app and wearable devices was examined, it could be seen that the traditional cardiology journals have given way to newer journals that focus on digital technology and medical informatics. The *Journal of Medical Internet Research* and JMIR-portfolio journals were found to be the major publishing venues for these papers. Therefore, readers should focus not only on traditional cardiology journals when seeking the latest advancements of digital technology use in the cardiology field.

XSL•FO
RenderX

There are several limitations. First, not all journals (and hence papers) are indexed by Web of Science. Alternative databases are available, each with their own shortcomings. For example, Scopus may contain erroneous data [127], Google Scholar does not allow automated extraction of title and abstract information, and PubMed does not contain citation data. Second, publication and citation counts do not necessarily equate to scientific quality. Within the diverse cardiology research field, the baseline research productivity in particular areas could be inhomogeneous; therefore, the ranking of clinicians or researchers is given for readers' general reference only. Notwithstanding, this study should allow readers to gain a better understanding of the literature on digital technology uses in cardiology.

Cardiovascular diseases that were frequently investigated in the literature included hypertension, atrial fibrillation, atherosclerosis, heart failure, and arterial stiffness. Recurring investigative modalities included ECG, angiography, echocardiography, digital subtraction angiography, and PPG. Readers searching for relevant information and authors searching for suitable publication venues for their work may consider that, while cardiology or cardiovascular system–focused journals were predominant in the overall literature set, the major publishing venues for the literature subset on smartphone apps and wearable devices were *Journal of Medical Internet Research* and JMIR-portfolio journals. Digital uses targeted physicians and patients as well as the general public, and their functions included assisting diagnosis, recording cardiovascular parameters, patient education, and teaching laypersons about cardiopulmonary resuscitation. The scientific body of literature on digital technology use in cardiology is rapidly growing, and its impact on health care is also expected to greatly increase in the near future.

Conflicts of Interest

None declared.

References

1. Kassal P, Steinberg MD, Steinberg IM. Wireless chemical sensors and biosensors: a review. Sens Actuators B Chem 2018 Aug;266:228-245. [doi: 10.1016/j.snb.2018.03.074]
2. Patel S, Park H, Bonato P, Chan L, Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil 2012 Apr 20;9(1):21 [FREE Full text] [doi: 10.1186/1743-0003-9-21] [Medline: 22520559]
3. Weiser M. The computer for the 21st century. Sci Am 1991 Sep;265(3):94-104. [doi: 10.1038/scientificamerican0991-94]
4. 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] [Medline: 31429410]
5. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017 Aug 03;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
6. Inkster B, Digital Mental Health Data Insights Group (DMHDIG). Early warning signs of a mental health tsunami: a coordinated response to gather initial data insights from multiple digital services providers. Front Digit Health 2020 Feb 10;2:578902 [FREE Full text] [doi: 10.3389/fdgth.2020.578902] [Medline: 34713053]
7. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
8. Bellini R, Wilson A, Smeddinck JD. Fragments of the past: curating peer support with perpetrators of domestic violence. 2021 May Presented at: Conference on Human Factors in Computing Systems; May 8-13, 2021; Yokohama. [doi: 10.1145/3411764.3445611]
9. Fortuna KL, Brooks JM, Umucu E, Walker R, Chow PI. Peer support: A human factor to enhance engagement in digital health behavior change interventions. J Technol Behav Sci 2019 Jun 15;4(2):152-161 [FREE Full text] [doi: 10.1007/s41347-019-00105-x] [Medline: 34337145]
10. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
11. Car J, Carlstedt-Duke J, Tudor Car L, Posadzki P, Whiting P, Zary N, Digital Health Education Collaboration. Digital education in health professions: the need for overarching evidence synthesis. J Med Internet Res 2019 Feb 14;21(2):e12913 [FREE Full text] [doi: 10.2196/12913] [Medline: 30762583]
12. Raisamo R, Rakkolainen I, Majaranta P, Salminen K, Rantala J, Farooq A. Human augmentation: past, present and future. Int J Hum Comput Stud 2019 Nov;131:131-143. [doi: 10.1016/j.ijhcs.2019.05.008]
13. Wang L, Lou Z, Jiang K, Shen G. Bio‐multifunctional smart wearable sensors for medical devices. Adv Intell Syst 2019 Aug 15;1(5):1900040. [doi: 10.1002/aisy.201900040]
14. Zhang H, Smeddinck J, Malaka R, Shu Y, Chen C, He B, et al. Wireless non-invasive motion tracking of functional behavior. Pervasive Mob Comput 2019 Mar;54:29-44. [doi: 10.1016/j.pmcj.2019.01.006]

XSL•FO
RenderX

15. Granados C, Pareja-Eastaway M. How do collaborative practices contribute to innovation in large organisations? the case of hackathons. Innovation 2019 Apr 05;21(4):487-505. [doi: 10.1080/14479338.2019.1585190]
16. Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Laxar D, Stoyanov J, et al. Virtual and augmented reality applications in medicine: analysis of the scientific literature. J Med Internet Res 2021 Feb 10;23(2):e25499 [FREE Full text] [doi: 10.2196/25499] [Medline: 33565986]
17. Bhavnani S, Parakh K, Atreja A, Druz R, Graham G, Hayek SS, et al. 2017 Roadmap for innovation-ACC health policy statement on health care transformation in the era of digital health, big data, and precision health: a report of the American College of Cardiology Task Force on Health Policy Statements and Systems of Care. J Am Coll Cardiol 2017 Nov 28;70(21):2696-2718 [FREE Full text] [doi: 10.1016/j.jacc.2017.10.018] [Medline: 29169478]
18. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. Big Data 2013 Jun;1(2):85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
19. Epstein DA, Ping A, Fogarty J, Munson SA. A lived informatics model of personal informatics. 2015 Presented at: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11, 2015; Osaka. [doi: 10.1145/2750858.2804250]
20. Hussein R. Medical informatics in the digital personalized health and medicine era: a SWOT analysis and actionable strategies. Stud Health Technol Inform 2020 Jun 16;270:869-873. [doi: 10.3233/SHTI200285] [Medline: 32570506]
21. Rooksby J, Rost M, Morrison A, Chalmers M. Personal tracking as lived informatics. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; May 1, 2014; Toronto. [doi: 10.1145/2556288.2557039]
22. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak 2008 Oct 06;8(1):45 [FREE Full text] [doi: 10.1186/1472-6947-8-45] [Medline: 18837999]
23. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
24. Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology - new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019 Nov 09;16(11):703-715 [FREE Full text] [doi: 10.1038/s41571-019-0252-y] [Medline: 31399699]
25. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artif Intell Med 2001 Aug;23(1):89-109. [doi: 10.1016/s0933-3657(01)00077-x]
26. Johnson AEW, Ghassemi MM, Nemati S, Niehaus KE, Clifton D, Clifford GD. Machine learning and decision support in critical care. Proc IEEE Inst Electr Electron Eng 2016 Feb;104(2):444-466 [FREE Full text] [doi: 10.1109/JPROC.2015.2501978] [Medline: 27765959]
27. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
28. Ashley EA. The precision medicine initiative: a new national effort. JAMA 2015 Jun 02;313(21):2119-2120. [doi: 10.1001/jama.2015.3595] [Medline: 25928209]
29. Bahcall O. Precision medicine. Nature 2015 Oct 15;526(7573):335-335. [doi: 10.1038/526335a] [Medline: 26469043]
30. Bzdok D, Varoquaux G, Steyerberg EW. Prediction, not association, paves the road to precision medicine. JAMA Psychiatry 2021 Feb 01;78(2):127-128. [doi: 10.1001/jamapsychiatry.2020.2549] [Medline: 32804995]
31. Malaka R, Herrlich M, Smeddinck J. Anticipation in motion-based games for health. In: Nadin M, editor. Anticipation and Medicine. Cham: Springer; 2017:351-363.
32. Tyler J, Choi SW, Tewari M. Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine. Curr Opin Syst Biol 2020 Apr;20:17-25 [FREE Full text] [doi: 10.1016/j.coisb.2020.07.001] [Medline: 32984661]
33. Swan M. Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen. J Pers Med 2012 Sep 12;2(3):93-118 [FREE Full text] [doi: 10.3390/jpm2030093] [Medline: 25562203]
34. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
35. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
36. Skjuve MB, Brandtzæg PB. Chatbots as a new user interface for providing health information to young people. In: Andersson Y, Dalquist U, Ohlsson J, editors. Youth and News in a Digital Media Environment: Nordic-Baltic Perspectives. Gothenburg: Nordicom; 2018.
37. Manero A, Smith P, Sparkman J, Dombrowski M, Courbin D, Kester A, et al. Implementation of 3D printing technology in the field of prosthetics: past, present, and future. Int J Environ Res Public Health 2019 May 10;16(9):1641 [FREE Full text] [doi: 10.3390/ijerph16091641] [Medline: 31083479]

XSL•FO
RenderX

38. Richterich A. When open source design is vital: critical making of DIY healthcare equipment during the COVID-19 pandemic. Health Sociol Rev 2020 Jul 10;29(2):158-167. [doi: 10.1080/14461242.2020.1784772] [Medline: 33411651]
39. Yetisen AK. Biohacking. Trends Biotechnol 2018 Aug;36(8):744-747. [doi: 10.1016/j.tibtech.2018.02.011] [Medline: 29550160]
40. Rehm HL. Evolving health care through personal genomics. Nat Rev Genet 2017 Apr 31;18(4):259-267 [FREE Full text] [doi: 10.1038/nrg.2016.162] [Medline: 28138143]
41. Kelman C, Bass A, Holman C. Research use of linked health data--a best practice protocol. Aust N Z J Public Health 2002;26(3):251-255. [doi: 10.1111/j.1467-842x.2002.tb00682.x] [Medline: 12141621]
42. Hodge JG, Gostin LO, Jacobson PD. Legal issues concerning electronic health information: privacy, quality, and liability. JAMA 1999 Oct 20;282(15):1466-1471. [doi: 10.1001/jama.282.15.1466] [Medline: 10535438]
43. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137 [FREE Full text] [doi: 10.1186/s12916-019-1377-7] [Medline: 31311535]
44. Anderson JG. Social, ethical and legal barriers to e-health. Int J Med Inform 2007;76(5-6):480-483. [doi: 10.1016/j.ijmedinf.2006.09.016] [Medline: 17064955]
45. Sankar PL, Parker LS. The Precision Medicine Initiative's All of Us research program: an agenda for research on its ethical, legal, and social issues. Genet Med 2017 Jul;19(7):743-750 [FREE Full text] [doi: 10.1038/gim.2016.183] [Medline: 27929525]
46. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
47. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
48. Smeddinck JD. Games for health. In: Dörner R, Göbel S, Kickmeier-Rust M, Masuch M, Zweig K, editors. Entertainment Computing and Serious Games. Cham: Springer; 2016:212-264.
49. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. Strategies to facilitate exposure to internet-delivered health behavior change interventions aimed at adolescents or young adults: a systematic review. Health Educ Behav 2011 Feb 28;38(1):49-62. [doi: 10.1177/1090198110372878] [Medline: 21189422]
50. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
51. 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] [Medline: 32877352]
52. Arigo D, Jake-Schoffman DE, Wolin K, Beckjord E, Hekler EB, Pagoto SL. The history and future of digital health in the field of behavioral medicine. J Behav Med 2019 Feb 1;42(1):67-83 [FREE Full text] [doi: 10.1007/s10865-018-9966-z] [Medline: 30825090]
53. Falter M, Scherrenberg M, Dendale P. Digital health in cardiac rehabilitation and secondary prevention: a search for the ideal tool. Sensors (Basel) 2020 Dec 22;21(1):12 [FREE Full text] [doi: 10.3390/s21010012] [Medline: 33374985]
54. Molinari G, Molinari M, Di Biase M, Brunetti ND. Telecardiology and its settings of application: an update. J Telemed Telecare 2017 Jan 13;24(5):373-381. [doi: 10.1177/1357633x16689432] [Medline: 28084886]
55. Scherrenberg M, Wilhelm M, Hansen D, Völler H, Cornelissen V, Frederix I, et al. The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2020 Jul 03;28(5):a. [doi: 10.1177/2047487320939671] [Medline: 33611532]
56. Dalal HM, Doherty P, McDonagh ST, Paul K, Taylor RS. Virtual and in-person cardiac rehabilitation. BMJ 2021 Jun 03;373:n1270 [FREE Full text] [doi: 10.1136/bmj.n1270] [Medline: 34083376]
57. Antoniades C, Asselbergs F, Vardas P. The year in cardiovascular medicine 2020: digital health and innovation. Eur Heart J 2021 Feb 14;42(7):732-739 [FREE Full text] [doi: 10.1093/eurheartj/ehaa1065] [Medline: 33388767]
58. Mesko B. A glimpse into the future of cardiology: how can advanced technologies lead to compassionate care? Eur Heart J 2021 Aug 31;42(33):3114-3116. [doi: 10.1093/eurheartj/ehab167] [Medline: 34038504]
59. Taj F, Klein MCA, van Halteren A. Digital health behavior change technology: bibliometric and scoping review of two decades of research. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13311 [FREE Full text] [doi: 10.2196/13311] [Medline: 31833836]
60. Gresham G, Schrack J, Gresham LM, Shinde AM, Hendifar AE, Tuli R, et al. Wearable activity monitors in oncology trials: current use of an emerging technology. Contemp Clin Trials 2018 Jan;64:13-21 [FREE Full text] [doi: 10.1016/j.cct.2017.11.002] [Medline: 29129704]
61. Singh N, Moneghetti K, Christle J, Hadley D, Froelicher V, Plews D. Heart rate variability: an old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. part two: prognosis and training. Arrhythm Electrophysiol Rev 2018 Dec;7(4):247-255 [FREE Full text] [doi: 10.15420/aer.2018.30.2] [Medline: 30588312]

XSL•FO
RenderX

62. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020 Jun 26;3(1):90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
63. Stappers NEH, Schipperijn J, Kremers SPJ, Bekker MPM, Jansen MWJ, de Vries NK, et al. Combining accelerometry and GPS to assess neighborhood-based physical activity: associations with perceived neighborhood walkability. Environ Behav 2020 Feb 17;53(7):732-752. [doi: 10.1177/0013916520906485]
64. Crutzen R. Hardwired… to self-destruct? Using technology to improve behavior change science. Health Psychol Bull 2021 May 21;5(1):70-80. [doi: 10.5334/hpb.26]
65. Wolfenden L, Nathan N, Williams CM. Computer-tailored interventions to facilitate health behavioural change. Br J Sports Med 2015 Nov 03;49(22):1478-1479. [doi: 10.1136/bjsports-2014-093508] [Medline: 24591065]
66. Perski O, Hébert ET, Naughton F, Hekler EB, Brown J, Businelle MS. Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: a systematic review. Addiction 2022 May 11;117(5):1220-1241. [doi: 10.1111/add.15687] [Medline: 34514668]
67. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun 2010 Aug 31;15 Suppl 2(sup2):9-19. [doi: 10.1080/10810730.2010.499985] [Medline: 20845189]
68. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health Informatics J 2016 Jun 10;22(2):81-98 [FREE Full text] [doi: 10.1177/1460458214534092] [Medline: 24916567]
69. Kickbusch I, Pelikan JM, Apfel F, Tsouros AD. Health literacy: the solid facts. World Health Organization Regional Office for Europe. 2013. URL: https://apps.who.int/iris/bitstream/handle/10665/128703/e96854.pdf [accessed 2022-04-22]
70. Prince L, Schmidtke C, Beck J, Hadden K. An assessment of organizational health literacy practices at an academic health center. Qual Manag Health Care 2018;27(2):93-97. [doi: 10.1097/QMH.0000000000000162] [Medline: 29596270]
71. Qi S, Hua F, Xu S, Zhou Z, Liu F. Trends of global health literacy research (1995–2020): analysis of mapping knowledge domains based on citation data mining. PLoS ONE 2021 Aug 9;16(8):e0254988. [doi: 10.1371/journal.pone.0254988]
72. 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] [Medline: 25843827]
73. Health Literacy: a Prescription to End Confusion. Washington DC: National Academies Press; 2004.
74. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
75. Manganello J, Gerstner G, Pergolino K, Graham Y, Falisi A, Strogatz D. The relationship of health literacy with use of digital technology for health information: implications for public health practice. J Public Health Manag Pract 2017;23(4):380-387. [doi: 10.1097/PHH.0000000000000366] [Medline: 26672402]
76. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
77. Wilcox L, Feiner S, Liu A, Restaino S, Collins S, Vawdrey D. Designing inpatient technology to meet the medication information needs of cardiology patients. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. 2012 Presented at: 2nd ACM SIGHIT International Health Informatics Symposium; January 28-30, 2012; Miami, Florida p. 831-836 URL: http://europepmc.org/abstract/MED/28018992 [doi: 10.1145/2110363.2110466]
78. Bickmore TW, Pfeifer LM, Jack BW. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. 2009 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 4-9, 2009; Boston, Massachusetts. [doi: 10.1145/1518701.1518891]
79. Seljelid B, Varsi C, Solberg Nes L, Stenehjem A, Bollerslev J, Børøsund E. Content and system development of a digital patient-provider communication tool to support shared decision making in chronic health care: InvolveMe. BMC Med Inform Decis Mak 2020 Mar 04;20(1):46 [FREE Full text] [doi: 10.1186/s12911-020-1065-8] [Medline: 32131808]
80. Sun S, Zhou X, Denny JC, Rosenbloom TS, Xu H. Messaging to your doctors: understanding patient-provider communications via a portal system. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France. [doi: 10.1145/2470654.2466230]
81. Magnani JW, Mujahid MS, Aronow HD, Cené CW, Dickson VV, Havranek E, American Heart Association Council on Epidemiology and Prevention; Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; Council on Peripheral Vascular Disease; Council on Quality of Care and Outcomes Research; Stroke Council. Health literacy and cardiovascular disease: fundamental relevance to primary and secondary prevention: a scientific statement from the American Heart Association. Circulation 2018 Jul 10;138(2):e48-e74 [FREE Full text] [doi: 10.1161/CIR.0000000000000579] [Medline: 29866648]
82. Yeung AWK, Souto EB, Durazzo A, Lucarini M, Novellino E, Tewari D, et al. Big impact of nanoparticles: analysis of the most cited nanopharmaceuticals and nanonutraceuticals research. Curr Res Biotechnol 2020 Nov;2:53-63. [doi: 10.1016/j.crbiot.2020.04.002]

XSL•FO
RenderX

83. Yeung AWK, Tzvetkov NT, Georgieva MG, Ognyanov IV, Kordos K, Jóźwik A, et al. Reactive oxygen species and their impact in neurodegenerative diseases: literature landscape analysis. Antioxid Redox Signal 2021 Feb 10;34(5):402-420. [doi: 10.1089/ars.2019.7952] [Medline: 32030995]
84. Yeung AWK, Tzvetkov NT, Jóźwik A, Horbanczuk OK, Polgar T, Pieczynska MD, et al. Food toxicology: quantitative analysis of the research field literature. Int J Food Sci Nutr 2020 Feb 29;71(1):13-21. [doi: 10.1080/09637486.2019.1620184] [Medline: 31140340]
85. Yeung AWK, Tzvetkov NT, El-Tawil OS, Bungău SG, Abdel-Daim MM, Atanasov AG. Antioxidants: scientific literature landscape analysis. Oxid Med Cell Longev 2019 Jan 08;2019:8278454-8278411 [FREE Full text] [doi: 10.1155/2019/8278454] [Medline: 30728893]
86. Durnin R, Glassner H, Jorgensen C, Fyler D. New approach to heart sound screening in children by analog-digital circuitry. Public Health Rep 1965 Sep;80(9):761-770 [FREE Full text] [Medline: 4953238]
87. Pemberton J, Li X, Karamlou T, Sandquist CA, Thiele K, Shen I, et al. The use of live three-dimensional Doppler echocardiography in the measurement of cardiac output: an in vivo animal study. J Am Coll Cardiol 2005 Feb 01;45(3):433-438 [FREE Full text] [doi: 10.1016/j.jacc.2004.10.046] [Medline: 15680724]
88. Pemberton J, Li X, Kenny A, Davies CH, Minette MS, Sahn DJ. Real-time 3-dimensional Doppler echocardiography for the assessment of stroke volume: an in vivo human study compared with standard 2-dimensional echocardiography. J Am Soc Echocardiogr 2005 Oct;18(10):1030-1036. [doi: 10.1016/j.echo.2005.03.009] [Medline: 16198879]
89. Shiota T, Jones M, Chikada M, Fleishman CE, Castellucci JB, Cotter B, et al. Real-time three-dimensional echocardiography for determining right ventricular stroke volume in an animal model of chronic right ventricular volume overload. Circulation 1998 May 19;97(19):1897-1900. [doi: 10.1161/01.cir.97.19.1897] [Medline: 9609081]
90. Mundt C, Montgomery K, Udoh U, Barker V, Thonier G, Tellier A, et al. A multiparameter wearable physiologic monitoring system for space and terrestrial applications. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):382-391. [doi: 10.1109/titb.2005.854509] [Medline: 16167692]
91. Liang Y, Chen Z, Ward R, Elgendi M. Hypertension assessment via ECG and PPG signals: an evaluation using MIMIC database. Diagnostics (Basel) 2018 Sep 10;8(3):65 [FREE Full text] [doi: 10.3390/diagnostics8030065] [Medline: 30201887]
92. Liang Y, Chen Z, Ward R, Elgendi M. Photoplethysmography and deep learning: enhancing hypertension risk stratification. Biosensors (Basel) 2018 Oct 26;8(4):101 [FREE Full text] [doi: 10.3390/bios8040101] [Medline: 30373211]
93. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759 [FREE Full text] [doi: 10.2196/21759] [Medline: 33074161]
94. Islam SMS, Maddison R. A comparison of blood pressure data obtained from wearable, ambulatory, and home blood pressure monitoring devices: prospective validation study. JMIR Data 2020;1(1):e22436 [FREE Full text] [doi: 10.2196/22436]
95. Bumgarner JM, Lambert CT, Hussein AA, Cantillon DJ, Baranowski B, Wolski K, et al. Smartwatch algorithm for automated detection of atrial fibrillation. J Am Coll Cardiol 2018 May 29;71(21):2381-2388 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.003] [Medline: 29535065]
96. William AD, Kanbour M, Callahan T, Bhargava M, Varma N, Rickard J, et al. Assessing the accuracy of an automated atrial fibrillation detection algorithm using smartphone technology: the iREAD study. Heart Rhythm 2018 Oct;15(10):1561-1565. [doi: 10.1016/j.hrthm.2018.06.037] [Medline: 30143448]
97. Kotecha D, Chua W, Fabritz L, Hendriks J, Casadei B, Schotten U, European Society of Cardiology (ESC) Atrial Fibrillation Guidelines Taskforce, the CATCH ME consortiumthe European Heart Rhythm Association (EHRA). European Society of Cardiology smartphone and tablet applications for patients with atrial fibrillation and their health care providers. Europace 2018 Feb 01;20(2):225-233 [FREE Full text] [doi: 10.1093/europace/eux299] [Medline: 29040548]
98. Curcio A, De Rosa S, Sabatino J, De Luca S, Bochicchio A, Polimeni A, et al. linical usefulness of a mobile application for the appropriate selection of the antiarrhythmic device in heart failure. Pacing Clin Electrophysiol 2016 Jul 19;39(7):696-702. [doi: 10.1111/pace.12872] [Medline: 27071370]
99. Salvi D, Poffley E, Orchard E, Tarassenko L. The mobile-based 6-minute walk test: usability study and algorithm development and validation. JMIR Mhealth Uhealth 2020 Jan 03;8(1):e13756 [FREE Full text] [doi: 10.2196/13756] [Medline: 31899457]
100. Wongvibulsin S, Habeos EE, Huynh PP, Xun H, Shan R, Porosnicu Rodriguez KA, et al. Digital health interventions for cardiac rehabilitation: systematic literature review. J Med Internet Res 2021 Feb 08;23(2):e18773 [FREE Full text] [doi: 10.2196/18773] [Medline: 33555259]
101. Creutzfeldt J, Hedman L, Heinrichs L, Youngblood P, Felländer-Tsai L. Cardiopulmonary resuscitation training in high school using avatars in virtual worlds: an international feasibility study. J Med Internet Res 2013 Jan 14;15(1):e9 [FREE Full text] [doi: 10.2196/jmir.1715] [Medline: 23318253]
102. Perron JE, Coffey MJ, Lovell-Simons A, Dominguez L, King ME, Ooi CY. Resuscitating cardiopulmonary resuscitation training in a virtual reality: prospective interventional study. J Med Internet Res 2021 Jul 29;23(7):e22920 [FREE Full text] [doi: 10.2196/22920] [Medline: 34326040]
103. Elgendi M. On the analysis of fingertip photoplethysmogram signals. Curr Cardiol Rev 2012 Feb 01;8(1):14-25 [FREE Full text] [doi: 10.2174/157340312801215782] [Medline: 22845812]

XSL•FO
RenderX

104. Kwon S, Kim H, Park KS. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. 2012 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 28-September 1. 2012; San Diego, California. [doi: 10.1109/embc.2012.6346392]
105. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021 Jul 01;18(7):465-478 [FREE Full text] [doi: 10.1038/s41569-020-00503-2] [Medline: 33526938]
106. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018 Jun 12;71(23):2668-2679 [FREE Full text] [doi: 10.1016/j.jacc.2018.03.521] [Medline: 29880128]
107. Paz O, Zhou X, Gillberg J, Tseng H, Gang E, Swerdlow C. Detection of T-wave alternans using an implantable cardioverter-defibrillator. Heart Rhythm 2006 Jul;3(7):791-797. [doi: 10.1016/j.hrthm.2006.03.022] [Medline: 16818208]
108. Porthan K, Viitasalo M, Toivonen L, Havulinna AS, Jula A, Tikkanen JT, et al. redictive value of electrocardiographic T-wave morphology parameters and T-wave peak to T-wave end interval for sudden cardiac death in the general population. Circ: Arrhythmia Electrophysiol 2013 Aug;6(4):690-696. [doi: 10.1161/circep.113.000356]
109. Yang C, Zhang W, Pang Z, Zhang J, Zou D, Zhang X, et al. A low-cost, ear-contactless electronic stethoscope powered by Raspberry Pi for auscultation of patients with COVID-19: prototype development and feasibility study. JMIR Med Inform 2021 Jan 19;9(1):e22753 [FREE Full text] [doi: 10.2196/22753] [Medline: 33436354]
110. Wali S, Guessi Margarido M, Shah A, Ware P, McDonald M, O'Sullivan M, et al. Expanding telemonitoring in a virtual world: a case study of the expansion of a heart failure telemonitoring program during the COVID-19 pandemic. J Med Internet Res 2021 Jan 22;23(1):e26165 [FREE Full text] [doi: 10.2196/26165] [Medline: 33444153]
111. Ahmadvand A, Kavanagh D, Clark M, Drennan J, Nissen L. Trends and visibility of "digital health" as a keyword in articles by JMIR publications in the new millennium: bibliographic-bibliometric analysis. J Med Internet Res 2019 Dec 19;21(12):e10477 [FREE Full text] [doi: 10.2196/10477] [Medline: 31855190]
112. Peng C, He M, Cutrona SL, Kiefe CI, Liu F, Wang Z. Theme trends and knowledge structure on mobile health apps: bibliometric analysis. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18212 [FREE Full text] [doi: 10.2196/18212] [Medline: 32716312]
113. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open 2020 Mar 02;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]
114. Cheng C, Beauchamp A, Elsworth GR, Osborne RH. Applying the electronic health literacy lens: systematic review of electronic health interventions targeted at socially disadvantaged groups. J Med Internet Res 2020 Aug 13;22(8):e18476 [FREE Full text] [doi: 10.2196/18476] [Medline: 32788144]
115. Walters R, Leslie SJ, Polson R, Cusack T, Gorely T. Establishing the efficacy of interventions to improve health literacy and health behaviours: a systematic review. BMC Public Health 2020 Jun 30;20(1):1040 [FREE Full text] [doi: 10.1186/s12889-020-08991-0] [Medline: 32605608]
116. Beauchamp A, Talevski J, Niebauer J, Gutenberg J, Kefalianos E, Mayr B, et al. Health literacy interventions for secondary prevention of coronary artery disease: a scoping review. Open Heart 2022 Jan;9(1) [FREE Full text] [doi: 10.1136/openhrt-2021-001895] [Medline: 35064057]
117. Brew-Sam N, Chib A. ow do smart device apps for diabetes self-management correspond with theoretical indicators of empowerment? an analysis of app features. Int J Technol Assess Health Care 2019 Apr 24;35(2):150-159. [doi: 10.1017/s0266462319000163]
118. Brew-Sam N, Chib A. How do smart device apps for diabetes self-management correspond with theoretical indicators of empowerment? an analysis of app features—CORRIGENDUM. Int J Technol Assess Health Care 2019 Jun 26;35(03):252. [doi: 10.1017/s0266462319000369]
119. Chow CK, Ariyarathna N, Islam SMS, Thiagalingam A, Redfern J. mHealth in cardiovascular health care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
120. MacKinnon GE, Brittain EL. Mobile health technologies in cardiopulmonary disease. Chest 2020 Mar;157(3):654-664 [FREE Full text] [doi: 10.1016/j.chest.2019.10.015] [Medline: 31678305]
121. Wongvibulsin S, Martin SS, Steinhubl SR, Muse ED. Connected health technology for cardiovascular disease prevention and management. Curr Treat Options Cardiovasc Med 2019 May 18;21(6):29 [FREE Full text] [doi: 10.1007/s11936-019-0729-0] [Medline: 31104157]
122. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
123. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Mar 03;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]

XSL•FO
RenderX

124. Meinhart F, Stütz T, Sareban M, Kulnik ST, Niebauer J. Mobile technologies to promote physical activity during cardiac rehabilitation: a scoping review. Sensors (Basel) 2020 Dec 24;21(1):65 [FREE Full text] [doi: 10.3390/s21010065] [Medline: 33374322]
125. Sarwar CM, Vaduganathan M, Anker SD, Coiro S, Papadimitriou L, Saltz J, et al. Mobile health applications in cardiovascular research. Int J Cardiol 2018 Oct 15;269:265-271. [doi: 10.1016/j.ijcard.2018.06.039] [Medline: 29921516]
126. Shuaib W, Khan MS, Shahid H, Valdes EA, Alweis R. Bibliometric analysis of the top 100 cited cardiovascular articles. Am J Cardiol 2015 Apr 01;115(7):972-981. [doi: 10.1016/j.amjcard.2015.01.029] [Medline: 25670637]
127. Franceschini F, Maisano D, Mastrogiacomo L. The museum of errors/horrors in Scopus. J Informetr 2016 Feb;10(1):174-182. [doi: 10.1016/j.joi.2015.11.006]

Abbreviations

ECG: electrocardiography
PPG: photoplethysmography

Edited by A Mavragani; submitted 31.12.21; peer-reviewed by I Madujibeya, N Maglaveras; comments to author 07.02.22; revised version received 14.03.22; accepted 15.03.22; published 11.05.22.

Please cite as:
Yeung AWK, Kulnik ST, Parvanov ED, Fassl A, Eibensteiner F, Völkl-Kernstock S, Kletecka-Pulker M, Crutzen R, Gutenberg J, Höppchen I, Niebauer J, Smeddinck JD, Willschke H, Atanasov AG
Research on Digital Technology Use in Cardiology: Bibliometric Analysis
J Med Internet Res 2022;24(5):e36086
URL: https://www.jmir.org/2022/5/e36086
doi:10.2196/36086
PMID:35544307

©Andy Wai Kan Yeung, Stefan Tino Kulnik, Emil D Parvanov, Anna Fassl, Fabian Eibensteiner, Sabine Völkl-Kernstock, Maria Kletecka-Pulker, Rik Crutzen, Johanna Gutenberg, Isabel Höppchen, Josef Niebauer, Jan David Smeddinck, Harald Willschke, Atanas G Atanasov. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.05.2022. 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.

XSL•FO
RenderX

Original Paper

Effect of a Popular Web Drama Video Series on HIV and Other Sexually Transmitted Infection Testing Among Gay, Bisexual, and Other Men Who Have Sex With Men in Singapore: Community-Based, Pragmatic, Randomized Controlled Trial

Rayner Kay Jin Tan[1,2], PhD; Wee Ling Koh[1], BSc; Daniel Le[3], BA; Sumita Banerjee[3], MSc; Martin Tze-Wei Chio[4], MSc; Roy Kum Wah Chan[5], MBBS; Christina Misa Wong[6], PhD; Bee Choo Tai[1], PhD; Mee Lian Wong[1], MBBS; Alex R Cook[1], PhD; Mark I-Cheng Chen[7], PhD; Chen Seong Wong[7], MBBS

[1]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

[2]University of North Carolina Project–China, Guangzhou, China

[3]Action for AIDS Singapore, Singapore, Singapore

[4]Department of Sexually Transmitted Infections Control Clinic, National Skin Centre, Singapore, Singapore

[5]National Skin Centre, Singapore, Singapore

[6]Behavioral, Epidemiological and Clinical Sciences, FHI 360, Durham, NC, United States

[7]National Centre for Infectious Diseases, Singapore, Singapore

Corresponding Author:
Rayner Kay Jin Tan, PhD
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
12 Science Drive 2
MD1 Tahir Foundation Building #10-01
Singapore, 117549
Singapore
Phone: 65 91878576
Email: rayner.tan@u.nus.edu

Abstract

Background: Gay, bisexual, and other men who have sex with men (GBMSM) are at disproportionately higher risk of acquiring HIV and other sexually transmitted infections (STI). While HIV/STI testing rates among GBMSM are increasing worldwide, they remain suboptimal in a variety of settings. While many studies have attempted to evaluate the efficacy of a variety of community-based campaigns, including peer and reminder-based interventions on HIV/STI testing, however few have attempted to do so for a web drama series.

Objective: This study evaluates the effectiveness of a popular web drama video series developed by a community-based organization in Singapore for GBMSM on HIV and other STI testing behaviors.

Methods: The study is a pragmatic, randomized controlled trial to evaluate a popular web drama video series developed by a community-based organization in Singapore for GBMSM. A total of 300 HIV-negative, GBMSM men in Singapore aged 18 to 29 years old were recruited and block-randomized into the intervention (n=150) and control arms (n=150). Primary outcomes included changes in self-reported intention to test for, actual testing for, and regularity of testing for HIV, syphilis, chlamydia or gonorrhea, while secondary outcomes include changes in a variety of other knowledge-based and psychosocial measures at the end of the study period.

Results: Overall, 83.3% (125/150) of participants in the intervention arm completed the proof of completion survey, compared to 88.7% (133/150) in the control arm. We found improvements in self-reporting as a regular (at least yearly) tester for HIV (15.9% difference, 95% CI, 3.2% to 28.6%; *P*=.02), as well as chlamydia or gonorrhea (15.5% difference, 95% CI, 4.2% to 26.9%; *P*=.009), indicating that the intervention had positively impacted these outcomes compared to the control condition. We also found improvements in participants' intentions to test for HIV (16.6% difference, 95% CI, 4.3% to 28.9%; *P*=.009), syphilis (14.8% difference, 95% CI, 3.2% to 26.4%; *P*=.01), as well as chlamydia or gonorrhea (15.4% difference, 95% CI, 4.2% to

XSL•FO
RenderX

26.6%; *P*=.008), in the next 3 months, indicating that the intervention was effective in positively impacting intention for HIV and other STI testing among participants.

Conclusions: There are clear benefits for promoting intentions to test regularly and prospectively on a broad scale through this intervention. This intervention also has potential to reach GBMSM who may not have access to conventional HIV and other STI prevention messaging, which have typically been implemented at sex-on-premises venues, bars, clubs, and in sexual health settings frequented by GBMSM. When coupled with community or population-wide structural interventions, the overall impact on testing will likely be significant.

Trial Registration: ClinicalTrials.gov NCT04021953; https://clinicaltrials.gov/ct2/show/NCT04021953

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-033855

(J Med Internet Res 2022;24(5):e31401) doi:10.2196/31401

KEYWORDS

HIV; STI; testing; health promotion; eHealth; mHealth

Introduction

Gay, bisexual, and other men who have sex with men (GBMSM) have been identified as key populations vulnerable to HIV acquisition [1]; however, rates of HIV testing have remained suboptimal among GBMSM in Southeast Asia. A study among young GBMSM in the Association of Southeast Asian Nations countries in 2015 found that 29.9% of participants had never had an HIV test, and these were more likely to be among younger GBMSM [2]. Unwillingness to know about their HIV status, fear of a positive result, fear of sexual orientation-related stigma or homophobia, and low perceived risk of HIV acquisition were factors found to be associated with lower rates of HIV testing among young GBMSM [3,4]. In Singapore, most individuals who test for HIV only do so through the course of medical care or through routine programmatic HIV screening, with only about 16.0% of the incident HIV cases in 2019 being diagnosed through voluntary screening. While a higher proportion (25.0%) of GBMSM tested through voluntary screening compared to heterosexual men (5.0%), diagnosis through voluntary screening remains suboptimal [5].

In general, numerous types of interventions exist that aim to increase HIV testing among GBMSM. These interventions include those that use aspects of peer education, outreach through social media, reminder-based systems, video-based interventions, and national social marketing campaigns. Social marketing campaigns have largely been promoted on a broader scale in non-Asian cities or settings where GBMSM reside [6-10], while reminder-based interventions have typically been implemented among GBMSM at sexual health clinics [11-13]. With the advent of geosocial networking smartphone apps, many interventions and campaigns now use key websites and mobile phone apps identified to be frequented by GBMSM for interventions as well [14-20].

Such interventions have reported varying degrees of effectiveness in achieving the aims of increasing HIV testing and overall disease awareness in Southeast Asia. This has been seen in the successes of the few social marketing campaigns on HIV and sexually transmitted infection (STI) testing in the region such as the “I Test, Do You?” campaign in Vietnam and the “TestXXX” campaigns in Vietnam, Thailand, the Philippines, and Indonesia [21,22].

There are, however, several limitations in the context of reach and feasibility for such interventions. For example, reminder-based and peer education-based interventions require existing health systems that can support such interventions, which may not be feasible in settings that do not have such services or where GBMSM-specific clinical services are unavailable due to the criminalization of sex between men. As such, these interventions may fall short of reaching out to more niche subsets of the GBMSM communities who may be more discreet about their sexual identities and hence may not often visit gay venues or sexual health clinics where these interventions are typically offered [23]. Furthermore, while social marketing campaigns have been effective in increasing the uptake of HIV/STI testing, such campaigns may not be feasible in settings such as Asia where negative perceptions of or attitudes toward GBMSM prevail [2].

A feasible option for interventions in such settings is the development of online, video-based interventions. However, evidence for the effectiveness of video-based interventions has not been conclusive. In a video-based intervention study conducted in Peru among GBMSM, differences in intention to test for HIV were not statistically significant between the intervention and control arm, although participants who identified as nongay did show increased willingness to do so [20]. Some studies have assessed the efficacy of crowdsourced videos on HIV testing and largely found that they were noninferior to regular health marketing campaigns [16] or had a positive effect on HIV testing rates through the use of home-based self-testing kits but not facility-based HIV/STI testing [24].

Given the gap in such research in Southeast Asia, this study sought to evaluate the effectiveness of a novel web drama series in achieving positive HIV/STI testing-related outcomes for young GBMSM. The videos used in the study form the second season of an educational and web drama miniseries, People Like Us, developed by gayhealth.sg and Action for AIDS (AFA) in 2018 [25]. The first season of the miniseries was screened at 10 film festivals and won several independent film awards. It also garnered more than 3 million views across various social media platforms since its launch in 2016. In spite of its popularity, little has been done to assess its effectiveness in positively impacting HIV and other STI testing-related

XSL•FO
RenderX

outcomes. Such popular online video interventions, which have been proven to be popular and easily accessible, may complement structural interventions and allow access to underserved or hard-to-reach subgroups of GBMSM.

Methods

Study Aims and Design

This is a pragmatic, parallel group, randomized controlled trial (RCT) to evaluate the efficacy of a web drama series developed by a community-based organization in Singapore in increasing an individual's intention to test, self-reported testing behaviors, and self-reported regularity of testing behaviors for HIV, syphilis, and other common STIs such as gonorrhea or chlamydia. The trial also aims to evaluate the impact of the web drama series on self-reported risk perception for HIV/STI; knowledge of risks associated with acquiring STIs and HIV; knowledge of HIV postexposure (PEP); knowledge of preexposure prophylaxis (PrEP); consistent condom use for anal sex with casual partners; incidence of STIs; connectedness to the lesbian, gay, bisexual, and transgender (LGBT) community; self-concealment of sexual orientation; perceived homophobia; internalized homophobia; HIV testing self-efficacy; and HIV testing social norms. The pragmatic nature of this trial arises due to the prospect of contamination, as the web drama series had already been launched in January 2019 and the trial was conducted among members of the community and subject to changes in the context. The implications of this are discussed later.

Inclusion Criteria

Inclusion criteria for participants in this study include self-reporting at the point of recruitment (1) an HIV-negative status or being unsure of one's HIV status; (2) being gay, bisexual, or queer in sexual orientation; (3) being of male gender, regardless of sex assigned at birth; (4) being aged 18 to 29 years; (5) being a Singapore citizen or permanent resident; (6) and having never watched an online video drama series by gayhealth.sg or AFA in the last year.

Ethics Approval

The study was registered at ClinicalTrials.gov (NCT04021953). Ethical approval was provided by the National University of Singapore institutional review board (reference: S-19-059).

Procedure and Randomization

Details of the intervention and the study procedures have been reported elsewhere in detail [26]. In brief, participants were recruited with the help of AFA and screened for eligibility through a short online survey. Throughout the entire survey process, personal identifiers were never directly linked to survey results, so as to protect the participants from potential criminal implications of disclosing their sexual activities with other men and other behaviors such as substance use. Upon completion of the enrollment survey and verification of eligibility, a staff member at AFA contacted eligible respondents to provide them with their participant ID number and formally invited them to participate in the study through the completion of the first online baseline survey. Respondents provided written consent for participation through an online participant information sheet prior to participating in the study. This survey was hosted on an encrypted, online survey administration website and took about 15 to 20 minutes to complete, and participants were reimbursed SGD 15.00 (US $10.84) for their time.

Upon completion of the baseline survey, participants were then randomly assigned in blocks of 6 in a 1:1 ratio to the intervention condition or control condition. Individuals who were assigned to the intervention condition were given a link to a series of 6 online videos, each about 10 minutes in duration, from the People Like Us web drama series, along with a link to an English-language online sexual health pamphlet tailored for GBMSM in Singapore. Individuals who were assigned to the control condition were scheduled to receive a link to the same online sexual health pamphlet as the standard of care for GBMSM at risk of acquiring HIV and other STIs in Singapore. All participants received their assigned conditions within 1 week after completing the baseline survey and were asked to complete a quiz 1 week after assignment to ascertain if participants had watched the online series of 6 videos or read the sexual health pamphlet. Participants received an SGD 20.00 (US $14.45) reimbursement following the completion of the quiz. Participants were not blinded to the group they have been assigned to and were told about their chances of being randomized to either group. At the 3-month and 6-month follow-up from the baseline, AFA contacted all eligible participants to continue with their follow-up surveys. Like the baseline survey, the second and third surveys were hosted on a survey administration website and took about 15 to 20 minutes to complete. Participants received SGD 15.00 (US $10.84) reimbursement for the completion of each survey.

Primary Outcome Measures

The survey questionnaire can be found in Multimedia Appendix 1. Primary outcomes for this evaluation included changes in self-reported intention to test for, ever testing for, testing in the last 6 months for, and regularity of testing for HIV, syphilis, and chlamydia or gonorrhea at the 6-month postintervention follow-up. For example, participants were asked "How likely are you to get tested for HIV in the next 3 months" to which they responded using a 6-point Likert scale from extremely unlikely to get tested to extremely likely to get tested. Self-reported testing was ascertained through the question "When did you go for your last (most recent) voluntary HIV test" (options to respond include never, in the last 3 months, in the last 6 months, 6 to 12 months ago, and more than 1 year ago), while self-reported regularity of testing was measured through the question "On average, how regularly do you test for HIV" (options to respond included I do not test regularly, once every few years, once a year, once every 6 months, once every 3 months, and once a month).

Secondary Outcome Measures

Secondary outcomes included changes in self-reported risk perception for HIV and other STIs, knowledge of HIV, knowledge of risks associated with acquiring other STIs, knowledge of HIV PEP and PrEP, self-reported consistent condom use for anal sex with casual partners (Cronbach α=.63, 95% CI, 0.62 to 0.65), self-reported incidence of STIs, and other

XSL•FO
RenderX

scales validated among GBMSM in other settings. These included scale measurements of connectedness to the LGBT community (Cronbach α=.87, 95% CI, 0.86 to 0.87) [27], self-concealment of sexual orientation (Cronbach α=.90, 95% CI, 0.90 to 0.91) [28], outness inventory (Cronbach α=.77, 95% CI, 0.77 to 0.79) [29], relevance of sexual orientation disclosure to sexual health care providers (Cronbach α=.83, 95% CI, 0.82 to 0.83), perceived homophobia (Cronbach α=.84, 95% CI, 0.83 to 0.84), internalized homophobia (Cronbach α=.85, 95% CI, 0.85 to 0.85) [30], HIV testing self-efficacy (Cronbach α=.90, 95% CI, 0.90 to 0.91) [31], and HIV testing social norms (Cronbach α=.55, 95% CI, 0.53 to 0.57) [32].

Sample Size

As the primary outcome of interest included HIV or other STI testing in the last 3 months, we used data from a previous study among 1098 GBMSM recruited through Grindr, the popular geosocial networking app [23,33]. The study found that 50.4% of respondents reported having had an HIV test in the 6 months prior to the survey. Assuming a 50% increase in recent HIV testing as a result of the intervention, as data from previous studies based on the impact such a web drama series on recent HIV testing remains limited [34], a sample size of 112 in each arm will yield statistical power higher than 80% to detect a significant change for the intervention based on calculations generated by a web-based sample size calculator (ClinCalc LLC) software. A target sample size of 150 participants per group was proposed to account for an attrition rate of 25% for each group across the 6-month follow-up.

Statistical Analyses

The baseline sociodemographic characteristics and primary outcome variables in the intervention and control groups were compared and any between-group differences were determined through chi-square tests. Intervention efficacy was analyzed over the entire study period (from baseline to the 6-month assessment) via chi-square tests for primary outcomes and Wilcoxon rank-sum tests for secondary outcomes with continuous variables, with median and interquartile ranges (IQR) also reported. Nonparametric tests were used for our analyses [35]. All analyses were evaluated based on the principle of intention to treat. While a 2-sided test at the 5% level of significance was indicated in our original protocol, we will discuss degrees of evidence in our paper instead [36]. We used the statistical software Stata (version 15, StataCorp LLC).

Results

Participant Characteristics

Overall, 777 participants were assessed for eligibility, and 482 participants were invited to participate to complete the baseline survey; 179 participants did not provide any contact details for follow-up, and 116 of them did not meet the eligibility criteria of being of male gender; identifying as gay, bisexual, or queer; and self-reporting as being HIV-negative. Overall, 83.3% (125/150) of participants in the intervention arm completed the proof of completion survey compared to 88.7% (133/150) in the control arm. At the first follow-up at 3 months, 8.0% (10/125) of participants in the intervention arm and 5.3% (7/133) in the control arm were lost to follow-up; at the second follow-up at 6 months, 5.2% (6/115) of participants in the intervention arm and 2.4% (3/126) in the control arm were lost to follow-up. Overall cumulative attrition rates reported for the intervention and control arms were 27.3% and 18.0%, respectively. The CONSORT (Consolidated Standards of Reporting Trials) diagram for the study is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Consolidated Standards of Reporting Trials diagram.

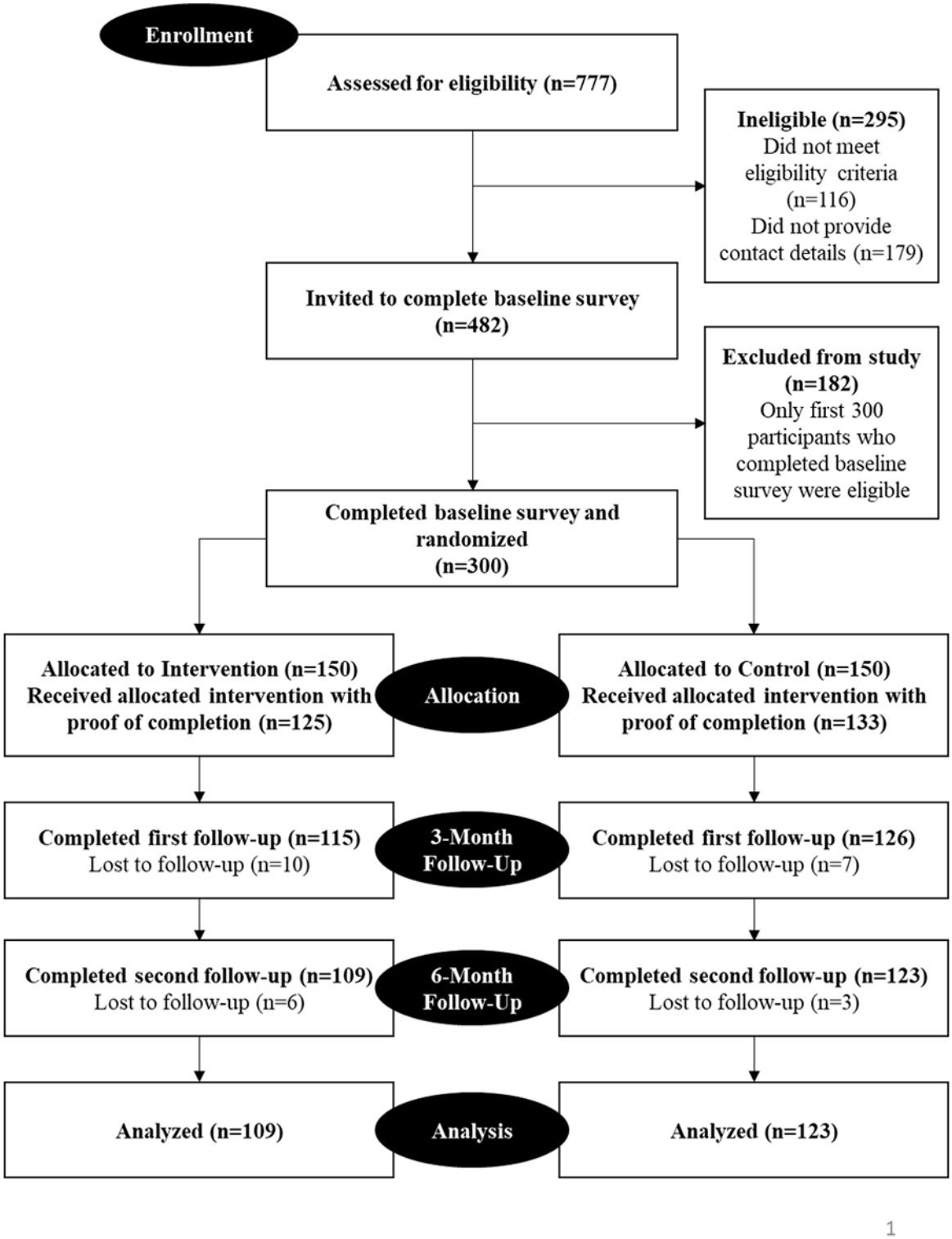


Table 1 summarizes patient demographic characteristics and outcome measures at baseline for the control and intervention groups. Overall, 150 participants were randomized to each arm. The 2 groups were generally well-balanced across all sociodemographic and baseline outcomes measures except for a slight imbalance for type of housing, where more participants in the intervention arm (126/150, 84.0%) reported staying in public housing relative to those in the control arm (112/150, 74.7%).

XSL•FO
RenderX

Table 1. Demographic characteristics and outcome variables for control and intervention groups.

Demographic and outcome variables at baseline	Control (n=150)	Intervention (n=150)
Demographic variables		
Age (year), mean (SD)	23.8 (2.99)	24.0 (2.98)
Chinese (ref: non-Chinese), n (%)	124 (82.7)	113 (75.3)
Gay (ref: bisexual, queer, or other), n (%)	127 (84.7)	116 (77.3)
Educational attainment below college (ref: some college), n (%)	95 (63.3)	92 (61.3)
Public housing (ref: private housing), n (%)	112 (74.7)	126 (84.0)
Monthly income below SGD[a] 5000 (ref: SGD 5000 and above), n (%)	82 (54.7)	73 (48.7)
Outcome variables		
Ever tested for HIV, n (%)	113 (75.3)	108 (72.0)
Ever tested for syphilis, n (%)	73 (48.7)	77 (51.3)
Ever tested for chlamydia or gonorrhea, n (%)	55 (36.7)	59 (39.3)
Tested for HIV in last 6 months, n (%)	66 (44.0)	63 (42.0)
Tested for syphilis in last 6 months, n (%)	39 (26.0)	41 (27.3)
Tested for chlamydia or gonorrhea in last 6 months, n (%)	21 (14.0)	26 (17.3)
Tested regularly (at least yearly) for HIV, n (%)	68 (45.3)	68 (45.3)
Tested regularly (at least yearly) for syphilis, n (%)	37 (24.7)	44 (29.3)
Tested regularly (at least yearly) for chlamydia or gonorrhea, n (%)	28 (18.7)	34 (22.7)
Intention to test for HIV in the next 3 months, n (%)	56 (37.3)	59 (39.3)
Intention to test for syphilis in the next 3 months, n (%)	41 (27.3)	39 (26.0)
Intention to test for chlamydia or gonorrhea in the next 3 months, n (%)	33 (22.0)	33 (22.0)

[a]SGD: Singapore dollar.

Primary Outcomes: HIV and Other STI Testing

We compared the primary outcomes for the intervention and control groups at 6 months postintervention, specifically, for ever testing, recent testing in the past 6 months, testing regularly (at least yearly), and intention to test in the next 3 months for HIV, syphilis, and chlamydia or gonorrhea. These findings are summarized in Table 2 and Figures 2 and 3. We observed marginal increases in recent testing for HIV, syphilis, and chlamydia or gonorrhea in the last 6 months. We also observed larger increases in the rates of ever testing for syphilis (11.3% difference, 95% CI, –1.4% to 24.0%) and for chlamydia and gonorrhea (9.9%, 95% CI, –2.9% to 22.7%).

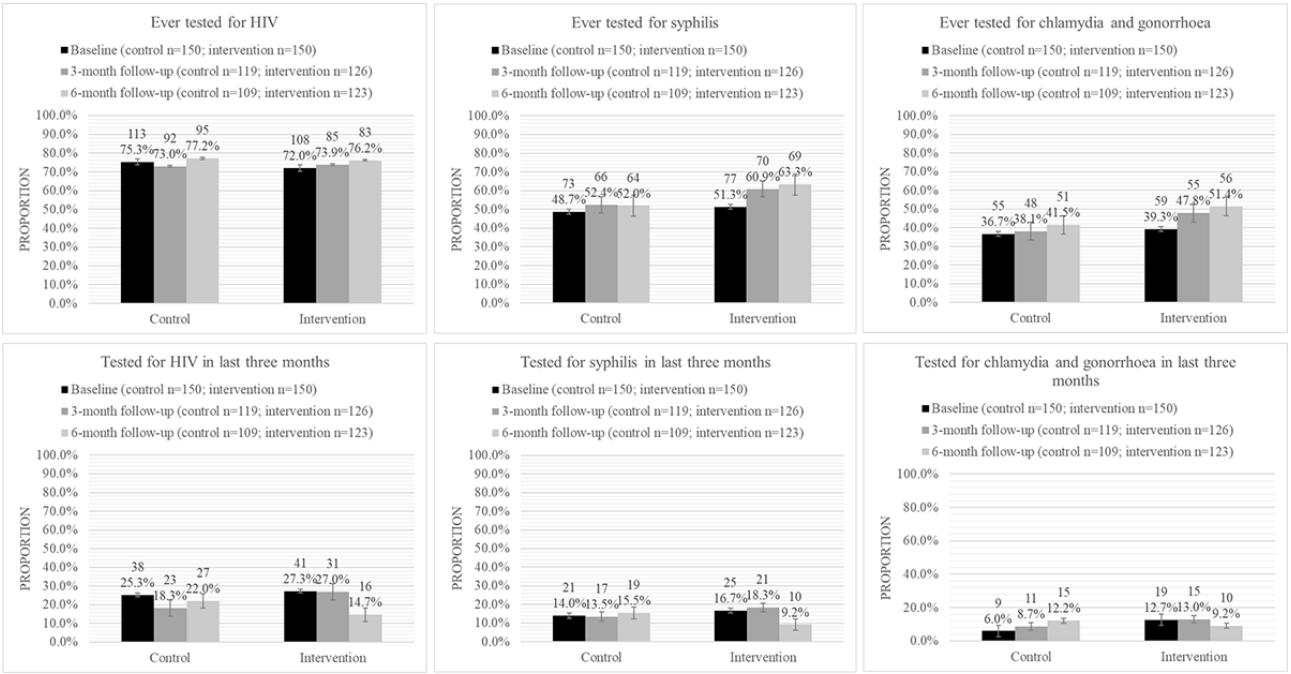
We found the greatest improvements in self-reporting for testing regularly (at least yearly) for HIV (15.9% difference, 95% CI, –28.6% to –3.2%) and for chlamydia or gonorrhea (15.5% difference, 95% CI, –26.9% to –4.2%), indicating that the intervention had positively impacted these outcomes compared to the control condition. We also found greatest improvements in participants' intentions to test for HIV (16.6% difference, 95% CI, –28.9% to –4.3%), syphilis (14.8% difference, 95% CI, –26.4% to –3.2%), and chlamydia or gonorrhea (15.4% difference, 95% CI, –26.6% to –4.2%) in the next 3 months, indicating that the intervention had positively impacted intention for HIV and other STI testing among participants in the intervention arm vis-à-vis those in the control arm.

XSL•FO
RenderX

Table 2. Comparison of primary outcome measures at 6 months postintervention.

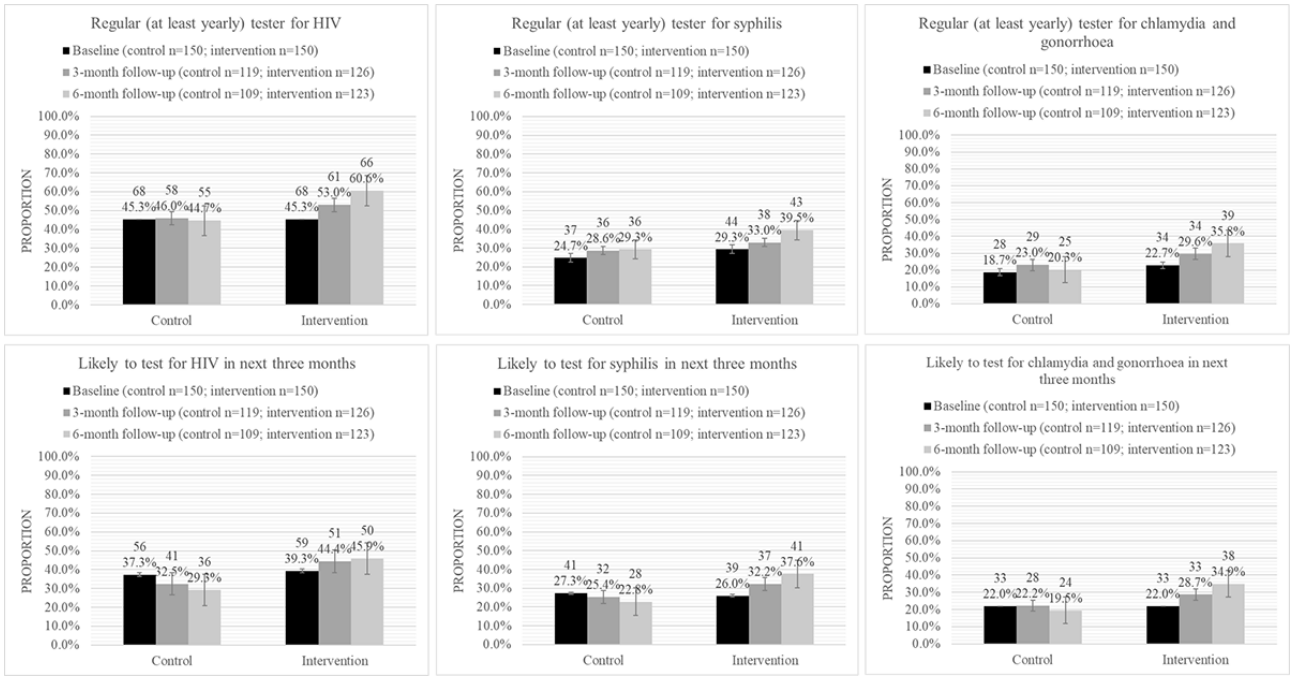
Primary outcome variables	Control (n=123), n (%)	Intervention (n=109), n (%)	Difference, %	95% CI	*P* value
Ever tested for HIV	95 (77.2)	83 (76.2)	–1.1	–11.9 to 9.9	.85
Ever tested for syphilis	64 (52.0)	69 (63.3)	11.3	–1.4 to 24.0	.08
Ever tested for chlamydia and gonorrhea	51 (41.5)	56 (51.4)	9.9	–2.9 to 22.7	.13
Tested for HIV in last 6 months	41 (33.3)	42 (38.5)	5.2	–7.1 to 17.5	.41
Tested for syphilis in last 6 months	31 (25.2)	30 (27.5)	2.3	–9.0 to 13.6	.69
Tested for chlamydia and gonorrhea in last 6 months	22 (17.9)	25 (22.9)	5.1	–5.3 to 15.3	.34
Tested regularly (at least yearly) for HIV	55 (44.7)	66 (60.6)	15.8	3.2 to 28.6	.02
Tested regularly (at least yearly) for syphilis	36 (29.3)	43 (39.5)	10.2	–2.0 to 22.4	.10
Tested regularly (at least yearly) for chlamydia and gonorrhea	25 (20.3)	39 (35.8)	15.5	4.2 to 26.9	.009
Intention to test for HIV in the next 3 months	36 (29.3)	50 (45.9)	16.6	4.3 to 28.9	.009
Intention to test for syphilis in the next 3 months	28 (22.8)	41 (37.6)	14.8	3.2 to 26.4	.01
Intention to test for chlamydia and gonorrhea in the next 3 months	24 (19.5)	38 (34.9)	15.4	4.2 to 26.6	.008

Figure 2. Trends in ever testing and recent testing for HIV and other sexually transmitted infections in control and intervention arms. High-resolution version of the figure is in Multimedia Appendix 2.



XSL•FO
RenderX

Figure 3. Trends in being a regular (at least yearly) tester and intention to test for HIV and other sexually transmitted infections in control and intervention arms. High-resolution version of the figure is in Multimedia Appendix 3.



Secondary Outcomes: HIV and Other STI Testing

We compared the secondary outcomes for the intervention and control groups at 6 months postintervention; specifically, perceived risk for HIV and other STIs, knowledge of HIV and other STIs, knowledge of HIV PEP and PrEP, inconsistent condom use for anal sex with casual partners in the last 6 months, recent STI diagnoses in the last 6 months, connectedness to the LGBT community, sexual orientation concealment, outness inventory, perceived relevance of sexual orientation disclosure in sexual health care, perceived homophobia, internalized homophobia, HIV testing self-efficacy, and HIV testing social norms. These findings are summarized in Table 3. Marginal differences in all secondary outcomes were observed.

XSL•FO
RenderX

Table 3. Comparison of secondary outcome measures at 6 months postintervention.

Secondary outcome variables	Control (n=123), median (IQR)	Intervention (n=109), median (IQR)	Difference[a], %	95% CI[a]	*P* value[b]
Perceived HIV risk, median (IQR)	2.0 (15.0)	6.0 (20.0)	0	0-0	.32
Perceived risk of other sexually transmitted infections, median (IQR)	5.0 (20.0)	8.0 (25.0)	0	0-0	.49
HIV and other sexually transmitted infections knowledge, median (IQR)	6.0 (2.0)	6.0 (1.0)	0	0-0	.08
Knowledge of HIV postexposure prophylaxis, n (%)	113 (91.9)	103 (94.5)	2.6	–3.9 to 9.1	.43
Knowledge of HIV preexposure prophylaxis, n (%)	122 (99.2)	107 (98.2)	–1.0	–3.9 to 1.9	.49
Inconsistent condom use in last 6 months with casual partners, n (%)	20 (16.3)	15 (13.8)	–2.5	–11.7 to 6.7	.60
Incidence of sexually transmitted infections in last 6 months, n (%)	8.0 (6.5)	4.0 (3.7)	–2.8	–8.4 to 2.8	.33
Connectedness to the lesbian, gay, bisexual, and transgender community, median (IQR)	22.0 (5.0)	21.0 (6.0)	1	–2.0 to 1.0	.35
Self-concealment of sexual orientation, median (IQR)	19.0 (10.0)	20.0 (9.0)	–1.0	–3.0 to 1.0	.30
Outness inventory, median (IQR)	2.74 (1.88)	2.33 (1.88)	0.1	–0.2 to 0.4	.45
Disclosure of sexual orientation to health care provider, median (IQR)	25.0 (8.0)	23.0 (9.0)	1	0 to 3.0	.17
Perceived homophobia, median (IQR)	18.0 (4.0)	18.0 (4.0)	0	–1.0 to 1.0	.94
Internalized homophobia, median (IQR)	10.0 (6.0)	10.0 (5.0)	0	–1.0 to 1.0	.56
HIV self-testing efficacy, median (IQR)	38.0 (11.0)	36.0 (11.0)	0	–2.0 to 2.0	.76
HIV testing social norms, median (IQR)	22.0 (4.0)	22.0 (4.0)	0	–1.0 to 0	.28

[a]Median differences between the groups and 95% CIs were estimated with the Hodges-Lehmann method.

[b]*P* value is derived from the Wilcoxon rank-sum test.

Discussion

Principal Findings

This study was a pragmatic, community-based, RCT on the effectiveness of the second season of the People Like Us web drama series. We found that the web drama series coupled with sexual health information provided through a pamphlet was more effective in promoting participants' self-reported intentions to test regularly (at least yearly) for chlamydia or gonorrhea and intentions to test for HIV and other STIs compared to the control condition that involved the availability of the pamphlet only. The intervention was developed to increase viewers' knowledge and perceptions of HIV and other STI risk, address homophobia and sexual orientation disclosure, increase safer-sex negotiation self-efficacy, promote positive attitudes toward condom use and other safe sex behaviors, build skills and self-efficacy for practicing safer sex, provide information on HIV and other STI testing and its benefits, provide information on resources for HIV and other STI testing and other mental health services, and model appropriate behaviors around practicing safer sex. As such, we hypothesized that the intervention would be more effective than traditional sexual health pamphlets in positively impacting a range of primary outcomes around HIV and other STI testing and secondary outcomes around HIV and other STI risk, safer sex, homophobia, HIV testing self-efficacy, and social norms.

The intervention was effective in promoting intentions to test for HIV and other STIs among participants in the intervention group compared to controls and, to a lesser degree, promoting ever testing for other STIs throughout the study period. We observed that recent testing for HIV and other STIs had also increased at the 3-month follow-up. We hypothesize that the lack of a strong impact of the intervention on actual testing variables may be attributed to the following reasons. The first reason, we hypothesize, may be attributed to the COVID-19 pandemic; the implementation of movement controls and legislation affecting social and sexual behaviors have caused an overall decrease in the rates of testing during the trial period. Given the descriptive increase in ever testing for other STIs and increase in recent testing for HIV and other STIs at the 3-month follow-up, there is reason to believe that this might have led to the marginal impact of the intervention for these variables. Second, while the intervention sought to potentially address individual beliefs around testing, it was not as effective in addressing complex psychosocial constructs such as perceived or internalized homophobia, self-efficacy, and social norms around testing, which were measured as secondary outcomes of the study and also serve as important constructs that underpin eventual testing behaviors [37-40]. These complex and often deep-rooted psychological constructs have been successfully modified in other types of more complex interventions that typically comprise components of training and engagement,

XSL•FO
RenderX

individual counselling, and peer engagement [41-43]. Finally, the lack of impact of the intervention on actual testing variables may be due to the impact of contamination in the study. At the end of the study, 13.8% (17/123) of participants in the control group had seen the People Like Us series and, in other words, had been exposed to the intervention prior to this study. As such, this may have biased the effect of the intervention toward the null. Unfortunately, the same question was not asked among participants who were in the intervention group, and we cannot ascertain if the dilution of the intervention effect might have occurred due to perhaps a high rate of exposure to the intervention prior to the study as well.

Pragmatic Nature of Trial

The People Like Us web drama series was launched in the community prior to the start of this study, and thus members of the community might have been exposed to the intervention prior to the study. Furthermore, researchers did not have an opportunity to participate in the development of the proposed intervention, and thus the intervention was not developed with a predetermined theory of behavior change. However, this study was designated to continue in view of its importance in the local context to evaluate the efficacy of such web drama series and justify further HIV and other STI prevention efforts that use online channels.

Given the pragmatic nature of the trial as described above, there was a possibility that control group participants may be exposed to the video series during the 6-month study period. To mitigate this, we ensured that details of the online video intervention (ie, title of web series, where to access it) were not included in the participant information sheet—only basic information on the possibility that they may be randomized to an online video intervention was mentioned. Furthermore, to reduce the possibility of contamination occurring in reaction to being asked the screening question, we avoided using the title of the web series but instead asked the question “Have you watched an online video drama series filmed by gayhealth.sg or AFA Singapore in the past year,” as this is gayhealth.sg and AFA’s only web series launched in the past year. While the generic nature of the question may have resulted in underreporting of viewing the video series, all participants eventually reported if they had viewed any of the episodes prior to or during the study period.

Specifically, participants in the treatment group were asked if they had previously watched any of the episodes when they submitted the intervention completion survey 1 week after the completion of their baseline survey, while the control group received a link to all 6 episodes of the video intervention alongside their final survey at the 6-month mark and were asked specifically which episodes they had watched prior to or during the intervention period.

The study team relied on self-reported outcomes such as testing behaviors and HIV/STI diagnoses as it is presently not possible to link clinic attendance or laboratory-confirmed diagnostic tests for HIV and other STIs to individual participants. These issues have arisen due to ethical concerns around linking participants’ personal information to survey results, which collect information on criminalized behavior such as sexual intercourse with other men, among participants in the sample. However, the findings of this proposed study would serve as a proof-of-concept for future studies that may be able to obtain funding and state support for other means of testing, such as the use of self-testing kits for HIV and other STIs.

Strengths and Limitations

The People Like Us web drama series has had vast reach among GBMSM in Singapore and around the world. As of April 2022, it has been nominated for the 48th International Emmy Awards for Best Short-Form Series and acquired by cable television and over-the-top media platforms such as HereTV, Gagaoolala, and Dekko and has attained more than 3 million views on YouTube. Given the viral nature of this web drama series, coupled with its vast reach among GBMSM, we believe that there are clear benefits to the promotion of such interventions in the GBMSM community. While it may not be able to address more complex constructs that underpin testing, there are clear benefits for promoting intentions to test regularly and prospectively for a wide audience; when coupled with community- or population-wide structural interventions, the overall impact on testing will be significant. Second, the nature of the web drama series allows this intervention to reach GBMSM who may not have access to conventional HIV and other STI prevention messaging that have typically been implemented at sex-on-premises venues, bars, clubs, and in sexual health settings frequented by GBMSM.

Limitations

We are also mindful of several limitations in this trial. First, given the pragmatic nature of the trial, we could not control for external situations such as the COVID-19 pandemic. As such, several aspects of the trial could not be evaluated in a manner that was intended and effects of the intervention may be underestimated in several respects. Second, contamination was a potential issue, where we found that at the end of the trial, 13.8% (17/123) of participants who remained in the control arm had seen the People Like Us season 2 series in spite of the initial screening question. However, intention-to-treat analysis was conducted, thus reducing the risk of overstating the effectiveness of the trial. Furthermore, given that most video-based interventions have been shown to largely influence short-term health behavior change rather than sustained, long-term behaviors [44], the potential decay and wash-out period for the intervention among these participants is less likely to have a substantial impact on the study results. Third, given the resource limitations of this study, we were not able to ascertain actual behaviors for HIV and other STI testing among participants through clinic attendance, and such measures were instead self-reported; however, we believe that the impact of any recall bias may have been minimal due to the introduction of the 3-month follow-up period between the pre- (baseline) and postintervention time points (6-month follow-up). Last, a formal process evaluation of the trial was not conducted due to a lack of resources, and thus we were not able to generate deeper insight into issues of implementation fidelity of the trial. Future trials should incorporate qualitative approaches to enhance our understanding of the individual and structural mechanisms that have led to effectiveness of the intervention among participants.

XSL•FO
RenderX

Nevertheless, this trial was conducted through a pilot study grant and will help to better inform future process evaluation efforts on larger trial studies similar in design. The pragmatic nature of the trial also meant that the trial was conducted in a community-based setting and subject to broader changes in the context in which it was rolled out. However, it was also during this time that COVID-19 started to take its hold as a pandemic, with the first case reported in Singapore on January 23, 2020. Further details on how this may have impacted the trial can be found in Multimedia Appendix 4 (see Figure S1 and Table S1 with explanatory notes).

Recommendations and Conclusions

Overall, this pragmatic, community-based RCT found that the critically acclaimed, wide-reaching People Like Us web drama series was effective in driving intentions to test regularly and prospectively for HIV and other STIs. We have several recommendations that accompany these findings. First, we recommend scaling up this intervention to a wider audience through further support and funding for marketing and promotional efforts, as this will help drive greater intentions to test among GBMSM, including those who may not be reached through traditional channels of sexual health communications and marketing. Second, we recommend rolling out additional programs alongside the marketing of the web drama series that address structural barriers to testing, such as issues of access and cost of testing, in the form of testing coupons or vouchers. Information on free and anonymous testing in the local GBMSM community by nongovernmental organizations such as AFA Singapore should be made available. Additionally, programs and workshops addressing deeply rooted, complex psychological constructs should also be offered to participants alongside the marketing of the web drama series to further drive HIV and other STI testing and related HIV prevention health behaviors. These may include workshops specifically on risks associated with HIV and other STIs, negotiating sexual relationships, and addressing homophobia and other topics underlying individual beliefs around HIV and other STIs that may underpin HIV and other STI testing behaviors.

Future trials and interventions should focus on addressing the limitations of this study. First, free clinic-based testing should be provided as outcomes for the trial to simultaneously address issues of structural barriers and limitations in outcome measurement. Second, the intervention should be structured to provide more complex, internet-based components such as online workshops, counselling, or peer support structures that may directly address the more complex, secondary outcomes of the study. Third, subgroup analyses to explore the impact of different demographic factors on intervention effectiveness can be conducted to further nuance and inform differentiated service delivery models targeting the rollout of such interventions for GBMSM. Finally, future studies should consider the potential impact that COVID-19 health promotion messaging, such as vaccinations and prevention methods, may have on HIV and other STI testing promotion in the community.

Acknowledgments

We would like to thank all who participated in this study. We also thank Action for AIDS Singapore and gayhealth.sg staff and volunteers who helped make this study possible. We are grateful for the team, cast, and crew who developed the People Like Us web drama series. We also thank the Saw Swee Hock School of Public Health and the Infectious Disease Program at the school for their support and funding of this study with grant SSHSPH ID-PRG/SeedFund/2018/03.

Authors' Contributions

RKJT, DL, SB, RKWC, and CSW conceptualized the study. MIC, ARC, RKWC, CMW, MLW, and CSW provided supervision for the study. RKJT acquired the funding for the study. RKJT, MIC, ARC, and BCT conducted formal analyses. RKJT, WLK, DL, SB, MTC. and CSW conducted the investigation and curated the data associated with the study. RKJT and WLK wrote the original draft. All coauthors reviewed and approved of the article prior to submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.
[DOCX File , 65 KB - jmir_v24i5e31401_app1.docx]

Multimedia Appendix 2
High-resolution version of Figure 2.
[PNG File , 765 KB - jmir_v24i5e31401_app2.png]

Multimedia Appendix 3
High-resolution version of Figure 3.
[PNG File , 823 KB - jmir_v24i5e31401_app3.png]

XSL•FO
RenderX

Multimedia Appendix 4
COVID-19 and its impact on the study.
[DOCX File , 122 KB - jmir_v24i5e31401_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1246 KB - jmir_v24i5e31401_app5.pdf]

References

1. Confronting inequalities: lessons for pandemic responses from 40 years of AIDS. UNAIDS. 2021. URL: https://www.unaids.org/sites/default/files/media_asset/2021-global-aids-update_en.pdf [accessed 2022-04-14]
2. Guadamuz TE, Cheung DH, Wei C, Koe S, Lim SH. Young, online and in the dark: scaling up HIV testing among MSM in ASEAN. PLoS One 2015 May 14;10(5):e0126658 [FREE Full text] [doi: 10.1371/journal.pone.0126658] [Medline: 25973907]
3. Scheim AI, Travers R. Barriers and facilitators to HIV and sexually transmitted infections testing for gay, bisexual, and other transgender men who have sex with men. AIDS Care 2017 Aug 27;29(8):990-995. [doi: 10.1080/09540121.2016.1271937] [Medline: 28027664]
4. Pharr JR, Lough NL, Ezeanolue EE. Barriers to HIV testing among young men who have sex with men (MSM): experiences from Clark County, Nevada. Glob J Health Sci 2015 Nov 03;8(7):9-17 [FREE Full text] [doi: 10.5539/gjhs.v8n7p9] [Medline: 26925893]
5. Update on the HIV/AIDS situation in Singapore 2019 (June 2020). Ministry of Health Singapore. URL: https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/hiv-stats/update-on-the-hiv-aids-situation-in-singapore-2019-(june-2020) [accessed 2020-09-24]
6. Stephens S, Bernstein K, McCright J, Klausner J. Dogs Are Talking: San Francisco's social marketing campaign to increase syphilis screening. Sex Transm Dis 2010 Mar;37(3):173-176. [doi: 10.1097/OLQ.0b013e3181bf5a80] [Medline: 19959974]
7. Plant A, Montoya JA, Rotblatt H, Kerndt PR, Mall KL, Pappas LG, et al. Stop the sores: the making and evaluation of a successful social marketing campaign. Health Promot Pract 2010 Jan 10;11(1):23-33. [doi: 10.1177/1524839907309376] [Medline: 18403747]
8. Plant A, Javanbakht M, Montoya J, Rotblatt H, O'Leary C, Kerndt P. Check Yourself: a social marketing campaign to increase syphilis screening in Los Angeles County. Sex Transm Dis 2014 Jan;41(1):50-57. [doi: 10.1097/OLQ.0000000000000069] [Medline: 24326583]
9. Pedrana A, Hellard M, Guy R, El-Hayek C, Gouillou M, Asselin J, et al. Stop the drama Downunder: a social marketing campaign increases HIV/sexually transmitted infection knowledge and testing in Australian gay men. Sex Transm Dis 2012 Aug;39(8):651-658. [doi: 10.1097/OLQ.0b013e318255df06] [Medline: 22801349]
10. Ahrens K, Kent CK, Montoya JA, Rotblatt H, McCright J, Kerndt P, et al. Healthy Penis: San Francisco's social marketing campaign to increase syphilis testing among gay and bisexual men. PLoS Med 2006 Dec 26;3(12):e474 [FREE Full text] [doi: 10.1371/journal.pmed.0030474] [Medline: 17194182]
11. Zou H, Meng X, Grulich A, Huang S, Jia T, Zhang X, et al. A randomised controlled trial to evaluate the impact of sexual health clinic based automated text message reminders on testing of HIV and other sexually transmitted infections in men who have sex with men in China: protocol for the T2T Study. BMJ Open 2017 Jul 10;7(7):e015787 [FREE Full text] [doi: 10.1136/bmjopen-2016-015787] [Medline: 28698334]
12. Zou H, Fairley CK, Guy R, Bilardi J, Bradshaw CS, Garland SM, et al. Automated, computer generated reminders and increased detection of gonorrhoea, chlamydia and syphilis in men who have sex with men. PLoS One 2013 Apr 17;8(4):e61972 [FREE Full text] [doi: 10.1371/journal.pone.0061972] [Medline: 23613989]
13. Bourne C, Knight V, Guy R, Wand H, Lu H, McNulty A. Short message service reminder intervention doubles sexually transmitted infection/HIV re-testing rates among men who have sex with men. Sex Transm Infect 2011 Apr 04;87(3):229-231. [doi: 10.1136/sti.2010.048397] [Medline: 21296796]
14. Zou H, Wu Z, Yu J, Li M, Ablimit M, Li F, et al. Internet-facilitated, voluntary counseling and testing (VCT) clinic-based HIV testing among men who have sex with men in China. PLoS One 2013 Feb 13;8(2):e51919 [FREE Full text] [doi: 10.1371/journal.pone.0051919] [Medline: 23418417]
15. Wang L, Podson D, Chen Z, Lu H, Wang V, Shepard C, et al. Using social media to increase HIV testing among men who have sex with men, Beijing, China, 2013-2017. MMWR Morb Mortal Wkly Rep 2019 May 31;68(21):478-482 [FREE Full text] [doi: 10.15585/mmwr.mm6821a3] [Medline: 31145721]
16. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV test promotion videos: a noninferiority randomized controlled trial in China. Clin Infect Dis 2016 Jun 01;62(11):1436-1442 [FREE Full text] [doi: 10.1093/cid/ciw171] [Medline: 27129465]

XSL•FO
RenderX

17. Lau JT, Tsui HY, Lau MM. A pilot clustered randomized control trial evaluating the efficacy of a network-based HIV peer-education intervention targeting men who have sex with men in Hong Kong, China. AIDS Care 2013 May 08;25(7):812-819. [doi: 10.1080/09540121.2012.749330] [Medline: 23244706]
18. Ko N, Hsieh C, Wang M, Lee C, Chen C, Chung A, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. J Med Internet Res 2013 Feb 25;15(2):e40 [FREE Full text] [doi: 10.2196/jmir.2264] [Medline: 23439583]
19. Klausner JD, Levine DK, Kent CK. Internet-based site-specific interventions for syphilis prevention among gay and bisexual men. AIDS Care 2004 Nov 27;16(8):964-970. [doi: 10.1080/09540120412331292471] [Medline: 15511728]
20. Blas MM, Alva IE, Carcamo CP, Cabello R, Goodreau SM, Kimball AM, et al. Effect of an online video-based intervention to increase HIV testing in men who have sex with men in Peru. PLoS One 2010 May 03;5(5):e10448 [FREE Full text] [doi: 10.1371/journal.pone.0010448] [Medline: 20454667]
21. APCOM's TestXXX initiative launches in Jakarta as TestJKT. Asia Pacific Coalition on Male Sexual Health. 2017. URL: https://www.apcom.org/apcoms-testxxx-initiative-launches-jakarta-testjkt/ [accessed 2020-01-08]
22. HIV self-testing in Vietnam. PATH. 2018. URL: https://www.path.org/resources/hiv-self-testing-in-vietnam/ [accessed 2022-04-14]
23. Tan RKJ, Teo AKJ, Kaur N, Harrison-Quintana J, Wong CS, Chen MI. Extent and selectivity of sexual orientation disclosure and its association with HIV and other STI testing patterns among gay, bisexual and other men who have sex with men. Sex Transm Infect 2019 Jun 27;95(4):273-278 [FREE Full text] [doi: 10.1136/sextrans-2018-053866] [Medline: 30918119]
24. Tang W, Wei C, Cao B, Wu D, Li KT, Lu H, et al. Crowdsourcing to expand HIV testing among men who have sex with men in China: a closed cohort stepped wedge cluster randomized controlled trial. PLoS Med 2018 Aug 28;15(8):e1002645 [FREE Full text] [doi: 10.1371/journal.pmed.1002645] [Medline: 30153265]
25. People Like Us. URL: https://www.gayhealth.sg/plu/ [accessed 2022-04-14]
26. Tan RKJ, Koh WL, Le D, Tan A, Tyler A, Tan C, et al. Effect of a web drama video series on HIV and other sexually transmitted infection testing among gay, bisexual and queer men: study protocol for a community-based, pragmatic randomised controlled trial in Singapore: the People Like Us (PLU) Evaluation Study. BMJ Open 2020 Apr 07;10(4):e033855 [FREE Full text] [doi: 10.1136/bmjopen-2019-033855] [Medline: 32269026]
27. Frost DM, Meyer IH. Measuring community connectedness among diverse sexual minority populations. J Sex Res 2012 Jan;49(1):36-49 [FREE Full text] [doi: 10.1080/00224499.2011.565427] [Medline: 21512945]
28. Schrimshaw EW, Siegel K, Downing MJ, Parsons JT. Disclosure and concealment of sexual orientation and the mental health of non-gay-identified, behaviorally bisexual men. J Consult Clin Psychol 2013 Feb;81(1):141-153 [FREE Full text] [doi: 10.1037/a0031272] [Medline: 23276123]
29. Mohr J, Fassinger R. Measuring dimensions of lesbian and gay male experience. Measurem Eval Counsel Devel 2019 Aug 29;33(2):66-90. [doi: 10.1080/07481756.2000.12068999]
30. Frost DM, Meyer IH. Internalized homophobia and relationship quality among lesbians, gay men, and bisexuals. J Couns Psychol 2009 Jan;56(1):97-109 [FREE Full text] [doi: 10.1037/a0012844] [Medline: 20047016]
31. Jamil MS, Guy RJ, Bavinton BR, Fairley CK, Grulich AE, Holt M, FORTH Investigator Group. HIV testing self-efficacy is associated with higher HIV testing frequency and perceived likelihood to self-test among gay and bisexual men. Sex Health 2017 Apr;14(2):170-178. [doi: 10.1071/SH16100] [Medline: 27883311]
32. Pettifor A, MacPhail C, Suchindran S, Delany-Moretlwe S. Factors associated with HIV testing among public sector clinic attendees in Johannesburg, South Africa. AIDS Behav 2010 Aug 18;14(4):913-921. [doi: 10.1007/s10461-008-9462-5] [Medline: 18931903]
33. Tan RKJ, Teo AKJ, Kaur N, Harrison-Quintana J, I-Cheng Chen M, Wong CS. Cost and anonymity as factors for the effective implementation of pre-exposure prophylaxis: an observational study among gay, bisexual and other men who have sex with men in Singapore. Sex Health 2018 Nov;15(6):533-541. [doi: 10.1071/SH18059] [Medline: 30249316]
34. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
35. Kitchen CM. Nonparametric vs parametric tests of location in biomedical research. Am J Ophthalmol 2009 Apr;147(4):571-572 [FREE Full text] [doi: 10.1016/j.ajo.2008.06.031] [Medline: 19327444]
36. Amrhein V, Korner-Nievergelt F, Roth T. The earth is flat (>0.05): significance thresholds and the crisis of unreplicable research. PeerJ 2017 Jul;5:e3544 [FREE Full text] [doi: 10.7717/peerj.3544] [Medline: 28698825]
37. Bakker AB, Buunk BP, Siero FW, van den Eijnden RJ. Application of a modified health belief model to hiv preventive behavioral intentions among gay and bisexual men. Psychol Health 1997 Aug;12(4):481-492. [doi: 10.1080/08870449708406724]
38. Felsher M, Szep Z, Krakower D, Martinez-Donate A, Tran N, Roth AM. "I don't need prep right now": a qualitative exploration of the barriers to PrEP care engagement through the application of the health belief model. AIDS Educ Prev 2018 Oct;30(5):369-381 [FREE Full text] [doi: 10.1521/aeap.2018.30.5.369] [Medline: 30332306]

XSL•FO
RenderX

39. Khumsaen N, Stephenson R. Beliefs and perception about HIV/AIDS, self-efficacy, and HIV sexual risk behaviors among young thai men who have sex with men. AIDS Educ Prev 2017 Apr;29(2):175-190 [FREE Full text] [doi: 10.1521/aeap.2017.29.2.175] [Medline: 28467158]
40. Wirtz AL, Kamba D, Jumbe V, Trapence G, Gubin R, Umar E, et al. A qualitative assessment of health seeking practices among and provision practices for men who have sex with men in Malawi. BMC Int Health Hum Rights 2014 Jun 03;14:20 [FREE Full text] [doi: 10.1186/1472-698X-14-20] [Medline: 24893654]
41. Hosek SG, Lemos D, Hotton AL, Fernandez MI, Telander K, Footer D, et al. An HIV intervention tailored for black young men who have sex with men in the House Ball Community. AIDS Care 2015 Oct 10;27(3):355-362 [FREE Full text] [doi: 10.1080/09540121.2014.963016] [Medline: 25300319]
42. Patel VV, Rawat S, Dange A, Lelutiu-Weinberger C, Golub SA. An internet-based, peer-delivered messaging intervention for HIV testing and condom use among men who have sex with men in India (CHALO!): pilot randomized comparative trial. JMIR Public Health Surveill 2020 Apr 16;6(2):e16494 [FREE Full text] [doi: 10.2196/16494] [Medline: 32297875]
43. Champenois K, Le Gall J, Jacquemin C, Jean S, Martin C, Rios L, et al. ANRS-COM'TEST: description of a community-based HIV testing intervention in non-medical settings for men who have sex with men. BMJ Open 2012 Apr 01;2(2):e000693 [FREE Full text] [doi: 10.1136/bmjopen-2011-000693] [Medline: 22466158]
44. 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 28;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]

Abbreviations

AFA: Action for AIDS
CONSORT: Consolidated Standards of Reporting Trials
GBMSM: gay, bisexual and other men who have sex with men
LGBT: lesbian, gay, bisexual, and transgender
PEP: HIV postexposure prophylaxis
PrEP: HIV preexposure prophylaxis
RCT: randomized controlled trial
SGD: Singapore dollar
STI: sexually transmitted infection

Edited by A Mavragani; submitted 20.06.21; peer-reviewed by A Rovetta, L Guo; comments to author 04.01.22; revised version received 09.03.22; accepted 09.04.22; published 06.05.22.

Please cite as:
Tan RKJ, Koh WL, Le D, Banerjee S, Chio MTW, Chan RKW, Wong CM, Tai BC, Wong ML, Cook AR, Chen MIC, Wong CS
Effect of a Popular Web Drama Video Series on HIV and Other Sexually Transmitted Infection Testing Among Gay, Bisexual, and Other Men Who Have Sex With Men in Singapore: Community-Based, Pragmatic, Randomized Controlled Trial
J Med Internet Res 2022;24(5):e31401
URL: https://www.jmir.org/2022/5/e31401
doi:10.2196/31401
PMID:35522470

©Rayner Kay Jin Tan, Wee Ling Koh, Daniel Le, Sumita Banerjee, Martin Tze-Wei Chio, Roy Kum Wah Chan, Christina Misa Wong, Bee Choo Tai, Mee Lian Wong, Alex R Cook, Mark I-Cheng Chen, Chen Seong Wong. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 06.05.2022. 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.

XSL•FO
RenderX

Original Paper

Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: 6-Month Follow-up Study of a Randomized Clinical Trial

Laura Garcia[1], PhD; Brandon Birckhead[2], MD; Parthasarathy Krishnamurthy[3], PhD; Ian Mackey[1], BA; Josh Sackman[1], MA; Vafi Salmasi[4], MD; Robert Louis[5], MD; Carina Castro[1], BA; Roselani Maddox[1], BSc; Todd Maddox[1], PhD; Beth D Darnall[6], PhD

[1]AppliedVR, Van Nuys, CA, United States
[2]Johns Hopkins School of Medicine, Baltimore, MD, United States
[3]University of Houston, Houston, TX, United States
[4]Stanford University School of Medicine, Palo Alto, CA, United States
[5]Hoag Memorial Hospital, Newport Beach, CA, United States
[6]Stanford School of Medicine, Palo Alto, CA, United States

Corresponding Author:
Todd Maddox, PhD
AppliedVR
16760 Stagg St
Suite 216
Van Nuys, CA, 91406
United States
Phone: 1 5129478494
Email: tmaddox@appliedvr.io

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2022/6/e40038

Abstract

Background: We previously reported the efficacy of an 8-week home-based therapeutic immersive virtual reality (VR) program in a double-blind randomized placebo-controlled study. Community-based adults with self-reported chronic low back pain were randomized 1:1 to receive either (1) a 56-day immersive therapeutic pain relief skills VR program (EaseVRx) or (2) a 56-day sham VR program. Immediate posttreatment results revealed the superiority of therapeutic VR over sham VR for reducing pain intensity; pain-related interference with activity, mood, and stress (but not sleep); physical function; and sleep disturbance. At 3 months posttreatment, therapeutic VR maintained superiority for reducing pain intensity and pain-related interference with activity, stress, and sleep (new finding).

Objective: This study assessed between-group and within-group treatment effects 6 months posttreatment to determine the extended efficacy, magnitude of efficacy, and clinical importance of home-based therapeutic VR.

Methods: E-surveys were deployed at pretreatment, end-of-treatment, and posttreatment months 1, 2, 3, and 6. Self-reported data for 188 participants were analyzed in a mixed-model framework using a marginal model to allow for correlated responses across the repeated measures. Primary outcomes were pain intensity and pain-related interference with activity, mood, stress, and sleep at 6 months posttreatment. Secondary outcomes were Patient-Reported Outcome Measurement Information System (PROMIS) sleep disturbance and physical function.

Results: Therapeutic VR maintained significant and clinically meaningful effects 6 months posttreatment and remained superior to sham VR for reducing pain intensity and pain-related interference with activity, stress, and sleep (d_s=0.44-0.54; P<.003). Between-group comparisons for physical function and sleep disturbance showed superiority of EaseVRx over sham VR (d_s=0.34; P=.02 and d_s=0.46; P<.001, respectively). Participants were encouraged to contact study staff with any problems experienced

XSL•FO
RenderX

during treatment; however, no participants contacted study staff to report adverse events of any type, including nausea and motion sickness.

Conclusions: Our 8-week home-based VR pain management program caused important reductions in pain intensity and interference up to 6 months after treatment. Additional studies are needed in diverse samples.

Trial Registration: ClinicalTrials.gov NCT04415177; https://clinicaltrials.gov/ct2/show/NCT04415177

International Registered Report Identifier (IRRID): RR2-10.2196/25291

(J Med Internet Res 2022;24(5):e37480) doi:10.2196/37480

KEYWORDS

behavioral health; chronic low back pain; treatment; virtual reality

Introduction

Chronic low back pain (CLBP) is the most common persistent pain condition worldwide, and multiple barriers impede patient access to timely and effective care. Innovations in digital therapeutics, such as immersive virtual reality (VR), offer the promise of home-based care, broad availability of treatment, and the potential to address the needs of underserved populations with CLBP.

Immersive VR is an evidence-based analgesic for acute low back pain [1], procedural low back pain [2], and CLBP [3,4]. Many VR treatments for CLBP involve rehabilitation exercise and require therapist guidance [5]. However, recent chronic pain research has investigated fully self-administered VR programs that require no clinician contact or guided movement exercises. Such programs closely mirror the content delivered in pain self-management or evidence-based psychological treatments for chronic pain.

In 2 randomized trials [3,4], we evaluated the effectiveness of a therapeutic VR program that incorporated multiple pain management modalities delivered via brief daily VR sessions. The first trial compared a 3-week skills-based VR program to the same therapeutic content delivered in audio-only format in 79 individuals with CLBP or fibromyalgia [3]. Posttreatment results revealed that the immersive VR modality was superior to the audio-only modality for reducing pain intensity and pain-related interference with activity, mood, sleep, and stress.

The second trial was a double-blind, randomized, placebo-controlled comparison of 8-week self-administered behavioral skills-based VR (EaseVRx; AppliedVR) with sham VR in 188 adults with CLBP. The 8-week sham VR program consisted of 2D placebo content involving nonimmersive nature scenes and neutral music (no skills training or pain education) [4,6]. Both treatments were delivered via the same commercial VR headsets and involved brief daily treatment sessions. Intention-to-treat analyses revealed benefits in both treatment groups and the superiority of therapeutic VR over sham VR for reducing pain intensity and pain-related interference with activity, stress, and mood, as well as sleep disturbance, with large effect sizes ranging from 1.17 to 1.3 (moderate to substantial clinical importance). On comparing the groups, a greater proportion of participants in the EaseVRx group achieved ≥30% reduction in pain intensity, and 46% of EaseVRx participants achieved ≥50% reduction in pain [4]. At 3 months, EaseVRx showed significant superiority over sham VR for reducing pain intensity and pain-related interference (activity, stress, and sleep [new finding]), with moderate to large effect sizes (0.56-0.88) exceeding the thresholds for clinical meaningfulness [7].

This study extended the results of this same study sample (N=188) [4,7] to 6 months posttreatment to evaluate further the durability of VR treatment. This study also included outcomes for participant blinding and treatment group unmasking at 6 months posttreatment. Finally, we investigated whether therapeutic VR engagement differs by socioeconomic status (SES), using a variable comprised of education level and annual household income.

Methods

Study Design

This 6-month follow-up study used a single-cohort, placebo-controlled, randomized clinical trial protocol [6]. The study involved an online national convenience sample of 188 community-based adults with self-reported CLBP.

The 6-month posttreatment data collection was completed in April 2021. Participants were instructed to return their VR headsets within 5 days of completing their 56-day treatment period (postage-paid packaging provided). Any headset returned after this 5-day shipment period was considered a late return.

This report contains participant-reported data from e-surveys deployed at pretreatment, end-of-treatment (day 56), and posttreatment months 1, 2, 3, and 6 for the primary outcomes (average pain intensity and pain-related interference with activity, mood, sleep, and stress) and the two secondary outcomes of sleep disturbance and physical function that demonstrated immediate effects after treatment.

Detailed information of the methods and interventions is provided in the study protocol [6].

Ethical Considerations

The Western Copernicus Group Institutional Review Board (Puyallup, WA) approved the study protocol in July 2020 (number: 1286465). Eligible individuals were enrolled after signing an eConsent form.

Participants

Individuals with CLBP were recruited nationally through Facebook and Google online advertisements, chronic pain organizations, and professional colleagues. Advertisements

XSL•FO
RenderX

directed individuals to the study website for information, and they were invited to complete an online eligibility form (see Textbox 1 for the inclusion/exclusion criteria). Figure 1 displays the participant study activities.

Textbox 1. Study inclusion and exclusion criteria.

Inclusion criteria

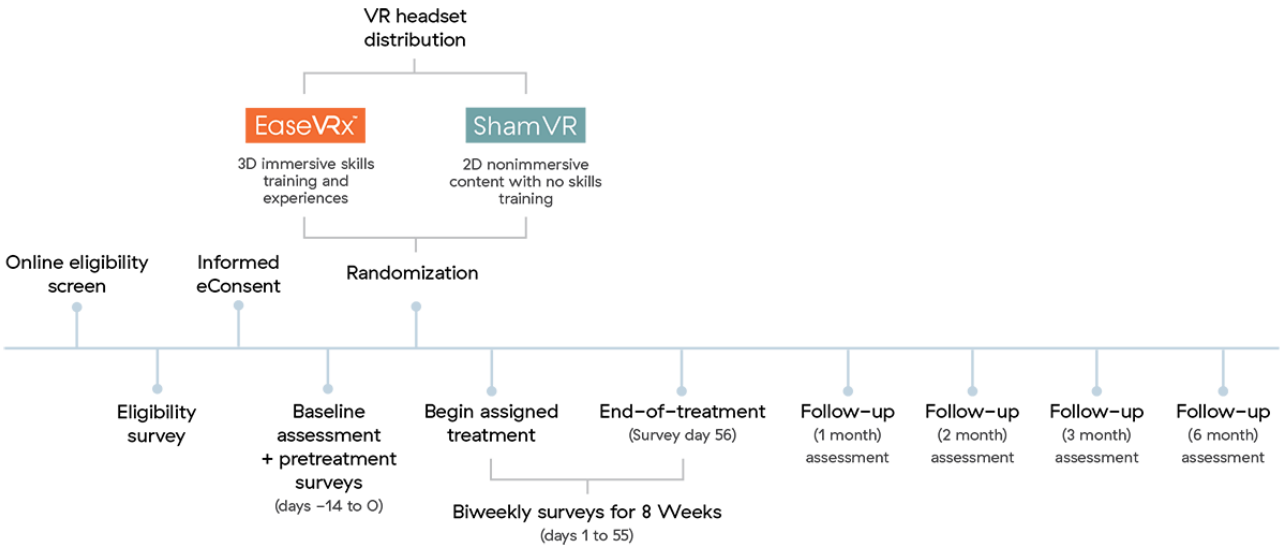
1. Men and women aged 18-85 years
2. Self-reported diagnosis of chronic low back pain without radicular symptoms
3. Chronic low back pain duration ≥6 months
4. Average pain intensity of ≥4 for the past month (0-10 numeric pain rating scale)
5. English fluency
6. Willing to comply with study procedures and restrictions
7. Wi-Fi access
8. Implicit de facto internet and computer literacy

Exclusion criteria

1. Gross cognitive impairment
2. Current or prior diagnosis of epilepsy, seizure disorder, dementia, migraines, or other neurological diseases that may prevent the use of virtual reality (VR) or predispose to adverse effects
3. Medical condition predisposing to nausea or dizziness
4. Hypersensitivity to flashing lights or motion
5. No stereoscopic vision or severe hearing impairment
6. Injury to the eyes, face, or neck that impedes comfortable use of a VR headset
7. Cancer-related pain
8. Depressive symptoms ≥2 on the Patient Health Questionnaire-2 (PHQ-2) depression screen
9. Previous use of EaseVRx for pain
10. Current or recent completion of participation (past 2 months) in any interventional research study
11. Currently pregnant or planning to become pregnant during the study period
12. Currently working at or having an immediate family member who works for a digital health company or pharmaceutical company that provides treatment for acute or chronic pain

XSL•FO
RenderX

Figure 1. Participant activities. VR: virtual reality.



Randomization and Participant Blinding

Enrolled participants completed a baseline survey battery and a pain survey that was readministered 3 times during the 2-week pretreatment period. These surveys were averaged to create a pretreatment pain value; completion of at least two surveys was required to progress to the treatment phase. The conduct of the study was entirely remote.

E-randomization was applied 1:1 without blocking and with participants allocated to either (1) a 56-day skills-based pain relief VR program (EaseVRx) or (2) a 56-day VR control condition (sham VR). Study participants understood they would be assigned to 1 of 2 VR treatments, but did not know that 1 treatment was sham. Participants remained blinded to their group assignment until all data were collected 6 months posttreatment. Participants were then informed that the study involved random assignment to VR with or without active treatment for chronic pain and were asked which program they believed they received. The statistician performed blinded analysis for the 56-day end-of-treatment results [4] and was unblinded to the individual group assignments for this study.

Procedures

All participants received a mailed Pico G2 4K all-in-one head-mounted VR device at no cost. The on-demand, easy-to-use, and commercially available Pico G2 4K device has a 3840×2160 screen, a 72 FPS frame rate, and minimal visual latency. Although the treatment content differed between the EaseVRx and sham VR devices, all packaging and directions were identical. Participants were given access to online instructional materials for their headset.

Participants were instructed to complete 1 VR program session daily for the treatment duration. Study staff monitored device use and sent reminders as needed for survey completion. At end-of-treatment, staff managed the postage-paid return of the devices. Posttreatment study staff interaction was limited to survey completion reminders and responses to participant inquiries.

Compensation included US $6 per survey during and after treatment ($150 possible; prorated; received as Amazon eGift cards). Participants who completed ≥16 study surveys during treatment were eligible to receive a VR headset after study completion (n=73).

Therapeutic VR (EaseVRx)

EaseVRx is a proprietary immersive, multimodal, skills-based, pain self-management VR program. EaseVRx incorporates evidence-based self-regulatory skills used in cognitive behavioral therapy for chronic pain (diaphragmatic breathing, biofeedback elements, cognition, and emotion regulation), mindfulness principles, and pain education into a multimodal therapeutic journey. The EaseVRx content is agnostic to pain type, condition, or disease (Figure 2).

XSL•FO
RenderX

Figure 2. Visual display of EaseVRx (skills-based, interactive, 3D) and sham VR (noninteractive, 2D nature scenes). VR: virtual reality.



The standardized 56-day program delivers the VR content through a prescribed sequence of daily immersive experiences grouped into 8 weekly themes relevant to living better with chronic pain. Content categories include pain education, relaxation and interoception, mindfulness escape, pain distraction games, and dynamic breathing. User exhalation is captured by an embedded microphone, providing interactive biodata-enabled therapeutics through synchrony with 3D visual displays and auditory feedback. VR sessions range from 2 to 16 minutes (average 6 minutes). Module content was designed to minimize emotional distress and cybersickness.

Sham VR

In compliance with VR-CORE clinical trial guidelines, we used an active and rigorous placebo comprised of nonimmersive 2D visual content [8]. Content included 20 rotating nature videos overlaid with music that was not relaxing, aversive, or distracting; content was devoid of pain education or pain management skills training (Figure 2). The average session duration closely matched that of EaseVRx.

Data Collection and Timepoints

Data were collected through REDCap Cloud for patient-reported outcomes at pretreatment, end-of-treatment, and posttreatment months 1, 2, 3, and 6.

The baseline survey included demographic variables and a battery of outcome measures. Demographic variables included age, gender, education level, race, ethnicity, employment status, annual household income, relationship status, back pain duration, state of residence, and zip code. The baseline survey included other measures of secondary outcomes that were omitted from our prior publication [7] and this report because we found no significant between-group changes for these measures at end-of-treatment when treatment effects are most pronounced (items from the Pain Catastrophizing Scale [9], 2-item Pain Self-Efficacy Questionnaire [10], and 8-item Chronic Pain Acceptance Questionnaire [11], and self-reported prescription opioid and over-the-counter analgesic medication use).

Measures

The Defense and Veterans Pain Rating Scale (DVPRS) [12] measured average pain intensity over the previous 24 hours using an 11-point numeric rating scale (0=no pain; 10=as bad as it could be and nothing else matters).

The DVPRS interference scale (DVPRS-II) [12] measured pain-related interference with activity, sleep, mood, and stress over the previous 24 hours (0=does not interfere; 10=completely interferes).

The National Institutes of Health (NIH) Patient-Reported Outcome Measurement Information System (PROMIS) short-form assessed physical function (version 6b) [13] and sleep disturbance (version 6a) [14] over the previous 7 days. The manuals' conversion tables were used to calculate individual short-form T scores using item response theory algorithms [15]. T scores were computed for individual response patterns using the Bayesian expected a posteriori method [15,16].

Adverse Event Monitoring

Participants were encouraged to contact staff about any problems with their device or treatment. Cybersickness was intended to be assessed immediately after treatment, but due to an error with the electronic survey, it was not captured until 1 month posttreatment.

Statistical Analyses

All analyses involved 2-sided hypothesis tests, with α=.05, and were adjusted for multiple comparisons within the family of tests as appropriate. Group equivalence was assessed through univariate tests of association between groups (EaseVRx/sham VR) for all baseline demographic and clinical variables, with the chi-square and Kruskal-Wallis tests applied as appropriate.

The intent-to-treat data were analyzed in a mixed-model framework (PROC GLIMMIX in SAS 9.4M6) using a marginal (population-averaged) model to allow for correlated responses across repeated measures. Explanatory factors included treatment group, time, and time × treatment group. Treatment group (EaseVRx/sham VR) was specified as a fixed-effects factor. Time (pretreatment, end-of-treatment, and posttreatment months 1, 2, 3, and 6) was specified as a random-effects factor

XSL•FO
RenderX

to allow for correlated responses using heterogeneous compound symmetry for the covariance structure within time. Analyses were conducted to assess (1) efficacy of treatment relative to pretreatment and (2) durability of treatment effects (end-of-treatment to month 6). Both analyses examined (1) EaseVRx vs sham VR between-group comparison across all timepoints and (2) whether the treatment group influenced the trajectory of the key variables over time. Efficacy, which included all 6 timepoints, was evidenced by significant treatment and time × treatment effects. We report multiplicity-adjusted Hochberg *P* values. Durability analyses were limited to end-of-treatment and posttreatment months 1, 2, 3, and 6. Durability was evidenced by a significant treatment effect but lack of time × treatment interaction, indicating sustained differences.

Missing values were not imputed for estimation of effects, but the predicted means were used in the graphical description. Linear mixed models were used as between-subject factors, and time of measurement was used as a within-subject factor. Effect sizes for the EaseVRx vs sham VR between-group comparison used the standardized mean difference version of Cohen *d* [17].

For each outcome variable, the effect size of the change pretreatment to 6 months posttreatment was assessed by treatment group using a repeated measures variation of Cohen *d* as d_{rm} owing to the within-subject nature of the comparison [17]. We applied common effect size thresholds of 0.3 (small), 0.5 (medium), and 0.8 (large). Clinical meaningfulness of the change in each outcome variable was further assessed by calculating the mean percent improvement from pretreatment to 6 months posttreatment and applying Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT)-recommended thresholds of magnitude for moderate (30%) and substantial (50%) clinical importance [18].

Participant blinding was assessed by the proportion of participants in each group who correctly determined their treatment assignment.

To test the feasibility of home-based VR in individuals with lower SES [19], we assessed therapeutic VR treatment engagement (total duration of treatment and number of sessions) in participants with lower SES (defined as ≤high school education or ≤US $59,999 median annual household income) vs higher SES (defined as >high school education or ≥US $60,000 median annual household income; US $60,000 was the selected threshold because it is below the US median household income of US $67,000) [20].

Results

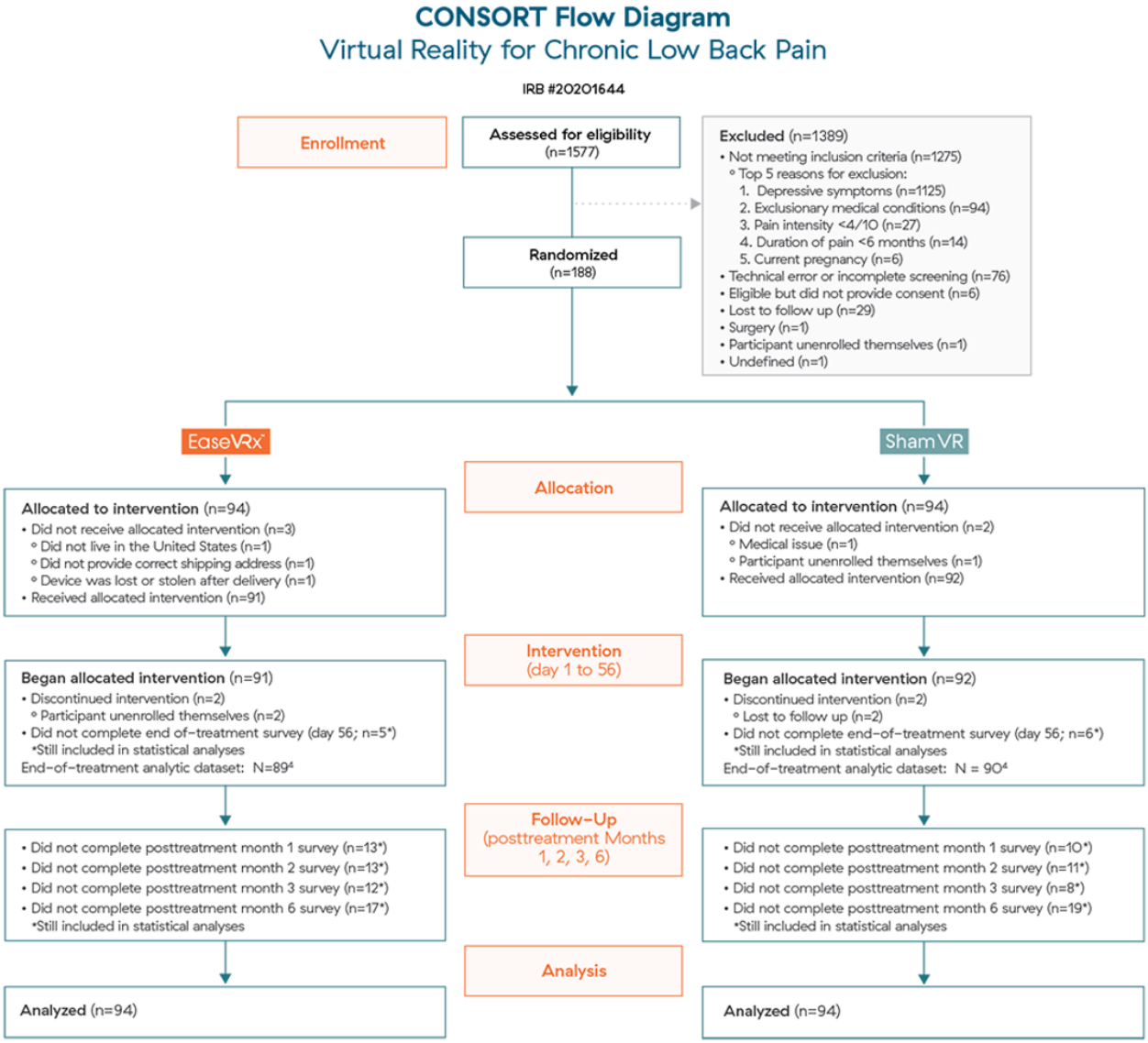
Overview

Recruitment took place from July 6, 2020, to July 30, 2020. Of 1577 individuals who completed an online eligibility screener, 1389 were excluded primarily for meeting or exceeding the threshold for depressive symptoms (see Figure 3 for the CONSORT diagram). In total, 188 individuals were enrolled, randomized, and allocated to a treatment group. All participants were included in the dataset regardless of treatment engagement or survey completion at posttreatment months 1, 2, 3, and 6. Previously reported device-use data revealed nonsignificant between-group differences for treatment engagement [4]. Posttreatment survey completion rates were 83% (n=156) for month 1, 82% (n=155) for month 2, 85% (n=159) for month 3, and 74% (n=139) for month 6.

Table 1 displays baseline demographics and clinical characteristics. Table 2 displays pretreatment outcome variables for each group. The sample included participants from 40 US states. The sample was predominantly female (145/188, 77.1%) and Caucasian (171/188, 91.0%), with most participants having at least some college education (171/188, 91.0%). The mean age was 51.7 years (SD 13.2 years; range 18-81 years), and the mean duration of CLBP was ≥5 years.

XSL•FO
RenderX

Figure 3. CONSORT flow diagram. VR: virtual reality.



XSL•FO
RenderX

Table 1. Baseline demographic and clinical characteristics by treatment group.

Variable	EaseVRx (n=94)	Sham VR[a] (n=94)
Gender, n (%)		
Male	23 (24)	20 (21)
Female	71 (76)	73 (78)
Other	0 (0)	1 (1)
Age (years), mean (SD)	52.1 (13.5)	51.3 (12.9)
Age range (years)	18.0-81.0	25.0-81.0
Age (years), median (IQR: Q1-Q3)	51.0 (41.0-62.0)	54.0 (41.0-62.0)
Race, n (%)		
Asian	2 (2)	1 (1)
Caucasian	82 (88)	88 (95)
African American	5 (5)	1 (1)
Multi-racial	2 (2)	3 (3)
Other	2 (2)	0 (0)
Missing	1 (1)	1 (1)
Education, n (%)		
High school graduate	6 (6)	10 (11)
Some college	22 (23)	17 (18)
Associate	10 (11)	16 (17)
Undergraduate	19 (20)	26 (28)
Postgraduate	37 (39)	24 (26)
Missing	0 (0)	1 (1)
Employment, n (%)		
Part time	9 (10)	7 (8)
Full time	39 (41)	36 (39)
Not working	13 (14)	11 (12)
Retired	17 (18)	21 (23)
Unable to work	16 (17)	18 (19)
Missing	0 (0)	1 (1)
Income, n (%)		
Less than US $40,000	25 (27)	24 (26)
US $40,000 to $59,999	24 (26)	19 (20)
US $60,000 to $79,999	16 (17)	19 (20)
Greater than US $80,000	28 (30)	32 (34)
Missing	1 (1)	0 (0)
Relationship, n (%)		
Married/civil union	55 (59)	63 (67)
Divorced/widowed/separated	21 (23)	14 (15)
Single	11 (12)	12 (13)
Single-cohabitating	6 (6)	5 (5)
Missing	1 (1)	0 (0)
Pain duration, n (%)		
<1 year	7 (7)	1 (1)

XSL•FO
RenderX

Variable	EaseVRx (n=94)	Sham VR[a] (n=94)
1 year to <5 years	25 (27)	26 (28)
5 years to <10 years	17 (18)	25 (27)
>10 years	45 (48)	42 (45)

[a]VR: virtual reality.

Table 2. Baseline outcome variables by treatment group.

Variable	EaseVRx (n=94)	Sham VR[a] (n=94)	*P* value[b]
Average pain intensity score			.61
Mean (SD)	5.1 (1.2)	5.2 (1.1)	
Range	2.2-8.2	2.8-8.0	
Median (IQR: Q1-Q3)	5.0 (4.2-5.8)	5.2 (4.4-5.8)	
Pain-related activity interference score			.43
Mean (SD)	5.3 (1.8)	5.5 (1.5)	
Range	1.2-10.0	1.0-8.8	
Median (IQR: Q1-Q3)	5.6 (4.0-6.6)	5.6 (4.6-6.3)	
Pain-related mood interference score			.27
Mean (SD)	4.4 (2.2)	4.7 (2.0)	
Range	0.0-8.8	0.2-9.6	
Median (IQR: Q1-Q3)	4.3 (2.8-5.8)	4.6 (3.4-6.0)	
Pain-related sleep interference score			.25
Mean (SD)	4.8 (2.6)	5.3 (1.9)	
Range	0.0-10.0	0.6-9.6	
Median (IQR: Q1-Q3)	5.0 (3.0-7.0)	5.4 (4.0-6.4)	
Pain-related stress interference score			.76
Mean (SD)	4.6 (2.2)	4.8 (2.0)	
Range	0.0-10.0	0.6-9.6	
Median (IQR: Q1-Q3)	4.7 (3.0-6.4)	5.0 (3.4-6.2)	
PROMIS[c] physical function score			.30
Mean (SD)	38.1 (5.1)	37.5 (4.7)	
Range	21.0-48.9	27.1-59.0	
Median (IQR: Q1-Q3)	37.6 (34.2-41.2)	37.6 (34.2-40.2)	
PROMIS sleep disturbance score			.17
Mean (SD)	56.7 (5.2)	57.7 (4.3)	
Range	44.2-67.5	45.5-69.0	
Median (IQR: Q1-Q3)	56.3 (53.3-60.4)	58.3 (55.3-60.4)	

[a]VR: virtual reality.

[b]Kruskal-Wallis *P* value.

[c]PROMIS: Patient-Reported Outcome Measurement Information System.

Primary Outcomes

We applied the analytic plan outlined above to each primary outcome. For each primary outcome figure referenced below, the x-axis represents time and the color bands represent 95% CI for the mean after correcting for multiple comparisons (Tukey-Kramer). Overlapping bands indicate nonsignificant group differences (*P* values) of simple main effects within each timepoint. Table 3 includes the corresponding model effects for each primary outcome in Figures 4-10.

XSL•FO
RenderX

Table 3. Model effects for primary outcomes.

Factor	Numerator df[a]	Denominator df	*F* value	*P* value
Pain intensity				
Treatment	1	186	11.05	.001
Time	5	758	47.43	<.001
Time × treatment	5	758	4.05	.001
Pain interference with activity				
Treatment	1	186	9.16	.003
Time	5	758	56.77	<.001
Time × treatment	5	758	2.95	.001
Pain interference with mood				
Treatment	1	186	10.59	.001
Time	5	758	35.66	<.001
Time × treatment	5	758	2.07	.07
Pain interference with sleep				
Treatment	1	186	13.82	<.001
Time	5	758	49.71	<.001
Time × treatment	5	758	1.84	.10
Pain interference with stress				
Treatment	1	186	10.23	.002
Time	5	758	46.94	<.001
Time × treatment	5	758	3.34	.006
PROMIS[b] physical function				
Treatment	1	186	5.57	.02
Time	5	758	22.78	<.001
Time × treatment	5	758	2.92	.01
PROMIS sleep disturbance				
Treatment	1	186	9.82	.002
Time	5	758	14.68	<.001
Time × treatment	5	758	3.78	.002

[a]df: degree of freedom.

[b]PROMIS: Patient-Reported Outcome Measurement Information System.

Pain Intensity

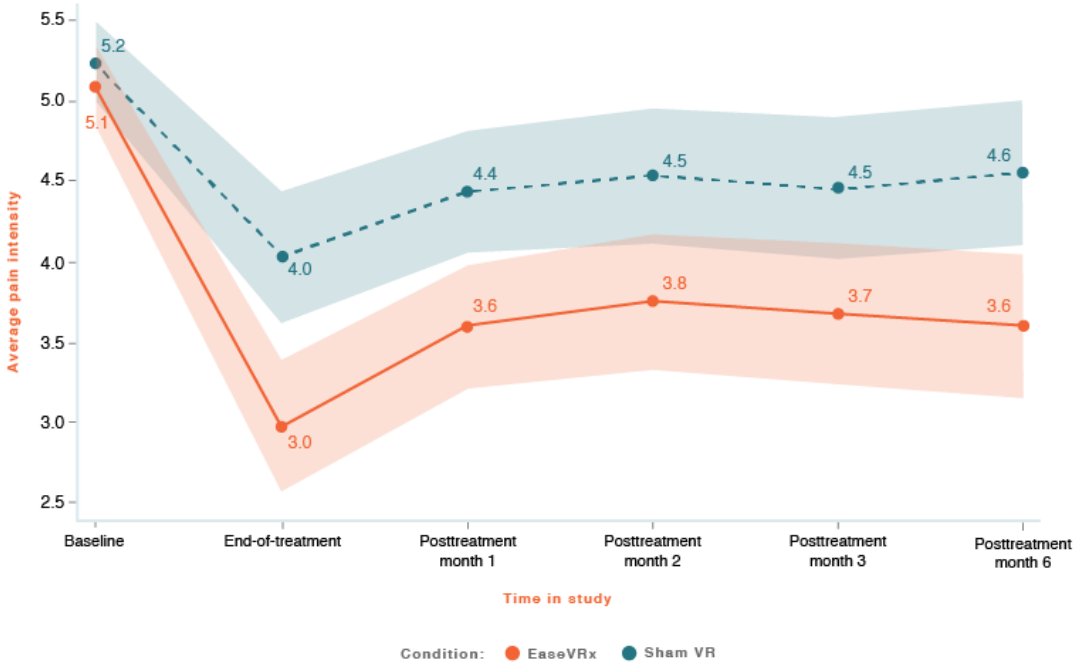
Average pain intensity was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.48; *P*=.001). Both treatment groups had lower average pain intensity from pretreatment to 6 months posttreatment (*P*<.001). While there was no between-group difference at pretreatment, at end-of-treatment, EaseVRx participants indicated lower pain intensity relative to sham VR, and this difference was maintained at month 6 (*P*=.001; Hochberg *P*=.006 after multiplicity correction; see Figure 4).

For pain intensity at 6 months posttreatment, the mean percentage change was −31.3% (moderate clinical importance) for the EaseVRx group and −15.9% (minimal clinical importance) for the sham VR group. We found that 52.1% (37/71) of EaseVRx and 25.0% (17/68) of sham VR participants achieved the threshold for moderate clinical meaningfulness (≥30%) and 38.0% (27/71) and 13.2% (9/68), respectively, achieved the threshold for substantial clinical meaningfulness (≥50%).

To evaluate durability, we compared end-of-treatment with the 6-month follow-up. On average, pain intensity was lower in the EaseVRx group than in the sham VR group (*P*=.004). We observed a significant effect of time (*P*<.001) but not time × treatment for pain intensity, indicating sustained superiority of EaseVRx over sham VR through 6 months posttreatment.

XSL•FO
RenderX

Figure 4. Average pain intensity. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. VR: virtual reality.



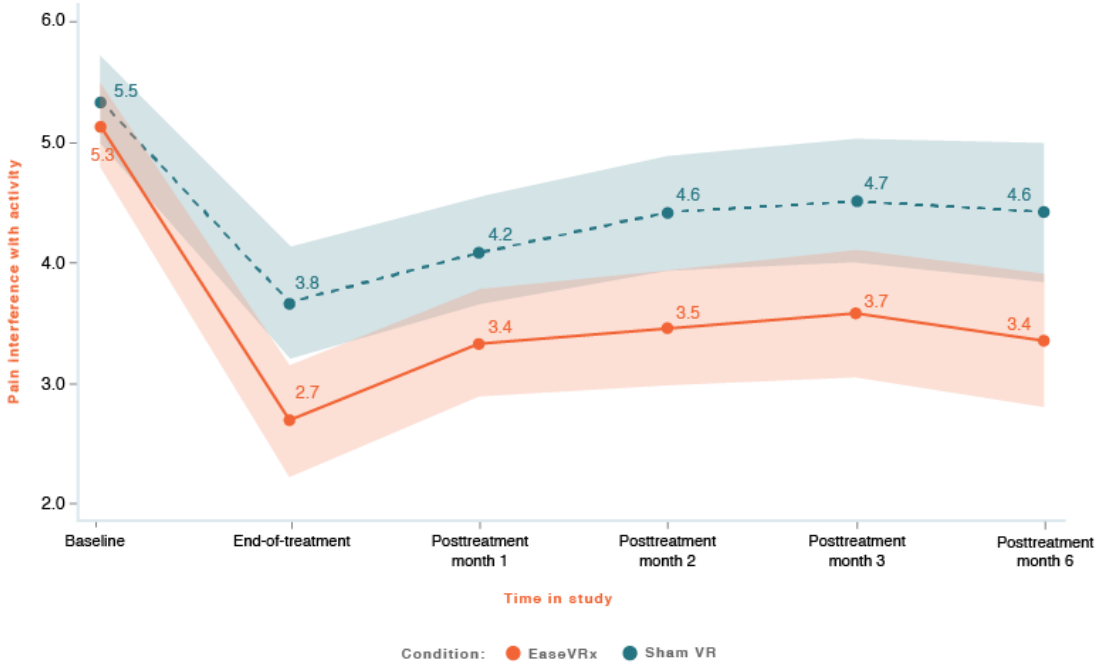
Pain-Related Interference With Activity

Average pain-related interference with activity was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.44; *P*=.003). Both groups had lower activity interference from pretreatment through month 6 (*P*<.001). Finally, we observed a pronounced between-group difference at end-of-treatment but not pretreatment (*P*=.01; Hochberg *P*=.04 after multiplicity correction; see Figure 5).

At 6 months posttreatment, the mean percentage change was −34.8% for the EaseVRx group and −20.8% for the sham VR group. We found that 60.6% (43/71) of EaseVRx and 39.7% (27/68) of sham VR participants achieved the threshold for moderate clinical meaningfulness and 50.7% (36/71) and 25.0% (17/68), respectively, achieved the threshold for substantial clinical meaningfulness.

Comparing end-of-treatment with the 6-month follow-up, pain-related interference with activity was lower in the EaseVRx group than in the sham VR group (*P*=.006). We observed a significant effect of time (*P*<.001) but not time × treatment (*P*=.92) for pain-related interference with activity, indicating sustained superiority of EaseVRx.

Figure 5. Pain interference with activity. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. VR: virtual reality.



XSL•FO
RenderX

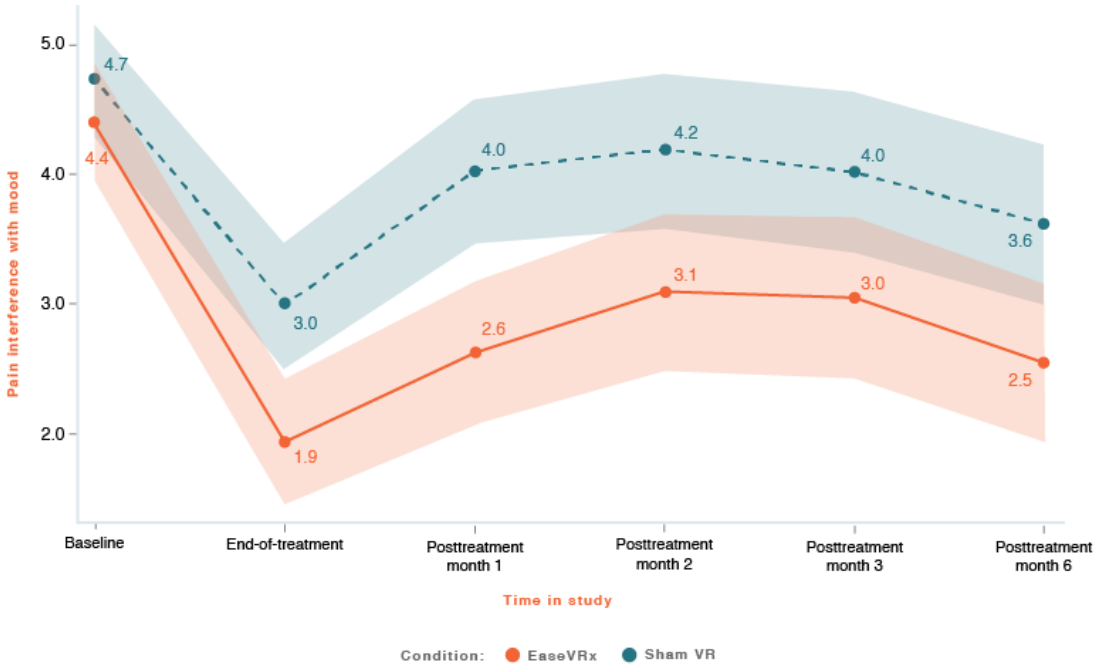
Pain-Related Interference With Mood

On average, pain-related interference with mood was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.47; *P*=.001). Both groups had lower mood interference from pretreatment through month 6 (*P*<.001). The time × treatment effect was not significant (*P*=.07; Hochberg *P*=.10 after multiplicity correction; see Figure 6).

At 6 months posttreatment, the mean percentage change for pain-related interference with mood was −39.2% for EaseVRx and −25.3% for sham VR. We found that 59.2% (42/71) of EaseVRx and 48.5% (33/68) of sham VR participants achieved the threshold for moderate clinical meaningfulness and 54.9% (39/71) and 41.2% (28/68), respectively, achieved the threshold for substantial clinical meaningfulness.

To evaluate durability, we compared end-of-treatment with the 6-month follow-up. On average, pain-related interference with mood was lower in the EaseVRx group than in the sham VR group (*P*=.003). We observed a significant effect of time (*P*<.001) but not time × treatment (*P*=.79), indicating sustained superiority of EaseVRx over sham VR through 6 months posttreatment.

Figure 6. Pain interference with mood. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. VR: virtual reality.



Pain-Related Interference With Sleep

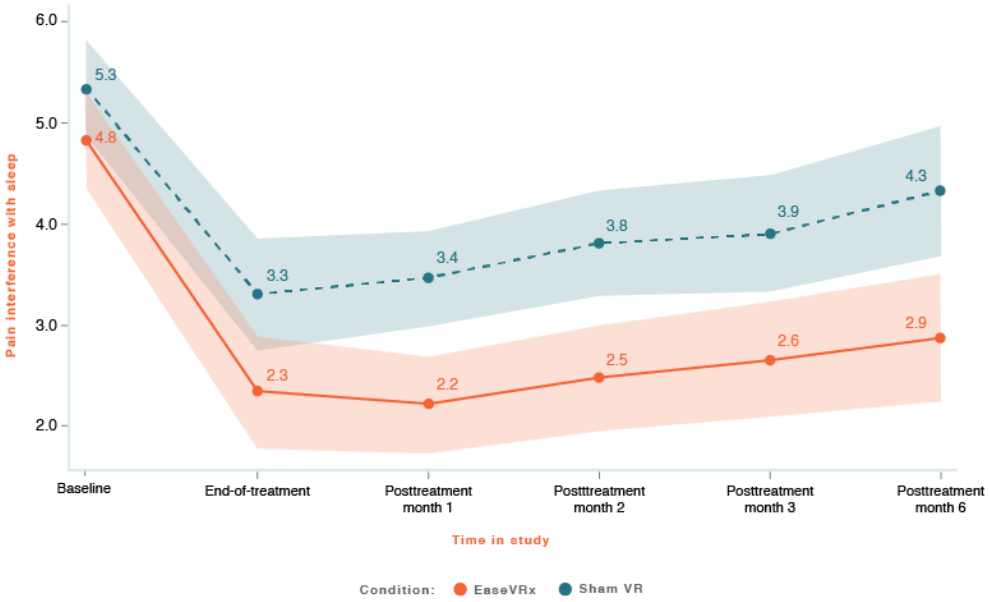
On average, pain-related sleep interference was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.54; *P*<.001). Both groups had lower sleep interference from pretreatment through month 6 (*P*<.001). The time × treatment effect was not significant (*P*=.10; Hochberg *P*=.10 after multiplicity correction; see Figure 7).

At 6 months posttreatment, the mean percentage change was −44.5% for the EaseVRx group and −18.9% for the sham VR group. We found that 63.4% (45/71) of EaseVRx and 45.6% (31/68) of sham VR participants achieved the threshold for moderate clinical meaningfulness and 47.9% (34/71) and 32.4% (22/68), respectively, achieved the threshold for substantial clinical meaningfulness.

Comparing end-of-treatment with the 6-month follow-up, pain-related interference with sleep was lower in the EaseVRx group than in the sham VR group (*P*<.001). We also observed a significant effect of time (*P*<.001) but not time × treatment (*P*=.89), indicating sustained superiority of EaseVRx over sham VR through 6 months posttreatment.

XSL•FO
RenderX

Figure 7. Pain interference with sleep. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. VR: virtual reality.



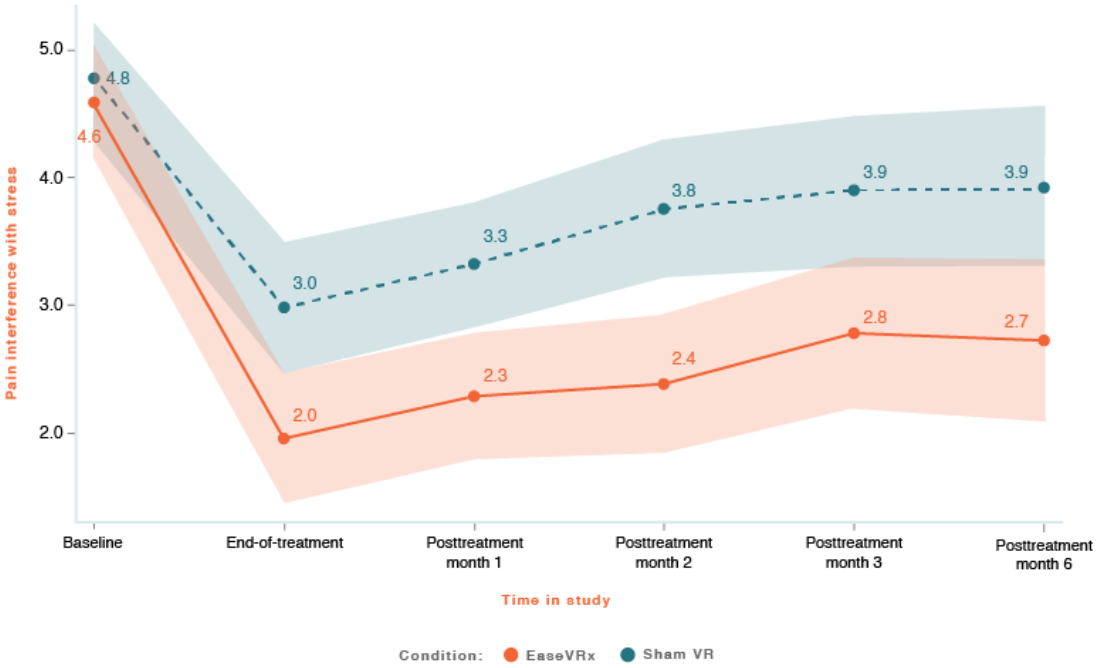
Pain-Related Interference With Stress

On average, pain-related stress interference was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.47; *P*=.001). Both groups had lower pain-related stress interference from pretreatment through month 6 (*P*<.001). While there was no between-group difference at pretreatment, there was a pronounced difference at end-of-treatment and at month 6 (time × treatment *P*=.006; Hochberg *P*=.02 after multiplicity correction; see Figure 8).

At 6 months posttreatment, the mean percentage change in pain-related interference with stress was −42.5% for the EaseVRx group and −23.3% for the sham VR group. We found that 67.6% (48/71) of EaseVRx and 39.7% (27/68) of sham VR participants achieved the threshold for moderate clinical meaningfulness and 60.6% (43/71) and 30.9% (21/68), respectively, achieved the threshold for substantial clinical meaningfulness.

Comparing end-of-treatment with the 6-month follow-up, pain-related interference with stress was lower in the EaseVRx group than in the sham VR group (*P*=.002). We observed a significant effect of time (*P*<.001) but not time × treatment (*P*=.86), indicating sustained superiority of EaseVRx over sham VR through 6 months posttreatment.

Figure 8. Pain interference with stress. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. VR: virtual reality.



XSL•FO
RenderX

Secondary Outcomes

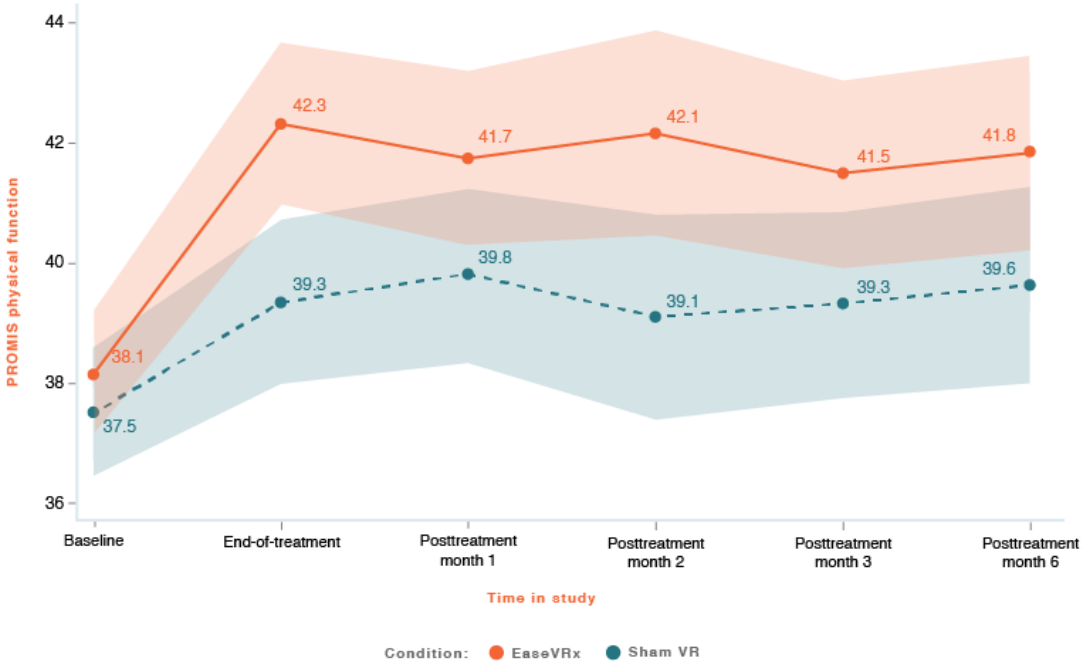
Physical Function

Physical function was higher in the EaseVRx group than in the sham VR group (Cohen d_s=0.34; *P*=.02), and both groups demonstrated increased physical function from pretreatment to 6 months posttreatment (*P*<.001). While there was no between-group difference at pretreatment, a between-group difference was pronounced at end-of-treatment through 6 months posttreatment (time × treatment *P*=.01; see Figure 9).

At 6 months posttreatment, the mean improvement in physical function was 10.5% for the EaseVRx group and 5.9% for the sham VR group, with changes in both groups categorized as clinically unimportant. We found that 12.7% (9/71) of EaseVRx and 4.4% (3/68) of sham VR participants reached the moderate clinical meaningfulness threshold. For substantial clinical meaningfulness, 4.2% (3/71) of EaseVRx participants and no sham VR participants achieved the threshold.

Comparing end-of-treatment with the 6-month follow-up, physical function was higher in the EaseVRx group than in the sham VR group (*P*=.02). The level of physical function was maintained from end-of-treatment to month 6 (*P*=.77). The time × treatment interaction effect was not significant (*P*=.45), indicating a sustained end-of-treatment effect (albeit of negligible clinical importance) for EaseVRx vs sham VR.

Figure 9. PROMIS physical function. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. PROMIS: Patient-Reported Outcome Measurement Information System; VR: virtual reality.



Sleep Disturbance

Sleep disturbance was lower in the EaseVRx group than in the sham VR group (Cohen d_s=0.46; *P*=.002). Both groups had decreased sleep disturbance over time (*P*<.001). While there was no between-group difference at pretreatment, at end-of-treatment, sleep disturbance was lower in the EaseVRx group than in the sham VR group, which did not sustain in posttreatment months 1, 2, 3, and 6 (time × treatment *P*=.002; see Figure 10).

At 6 months posttreatment, the mean percentage change in sleep disturbance was −8.8% for EaseVRx and −2.1% for sham VR. While 8.5% (6/71) of EaseVRx and 1.5% (1/68) of sham VR participants achieved the moderate clinical meaningfulness threshold, no EaseVRx or sham VR participants achieved the threshold for substantial clinical meaningfulness.

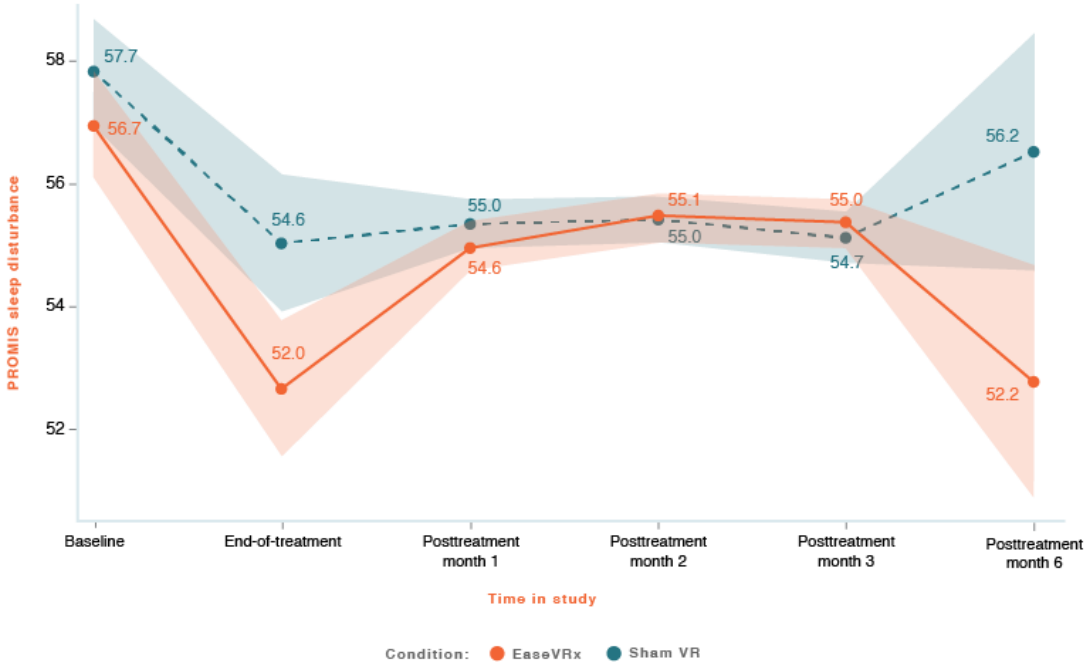
Comparing end-of-treatment with 6 months post-treatment, sleep disturbance was lower in the EaseVRx group than in the sham VR group (*P*=.002). There was a significant effect of time (*P*=.003). The end-of-treatment superiority of EaseVRx over sham VR for reduction in sleep disturbance was absent in posttreatment months 1, 2, and 3, and re-emerged at month 6 (time × treatment *P*=.002).

Assessing participant blinding 6 months posttreatment, 75% of EaseVRx and 71% of sham VR participants accurately identified their randomly assigned treatment. These proportions did not differ between groups (*P*>.05) and were significantly above chance.

Lower SES (n=97) and higher SES (n=91) participants were statistically equivalent for treatment engagement as indexed by the total duration of EaseVRx treatment time and total number of EaseVRx experiences.

XSL•FO
RenderX

Figure 10. PROMIS sleep disturbance. The color bands represent 95% CI for the mean after correcting for multiple comparisons. Overlapping bands indicate nonsignificant group differences of simple main effects within each timepoint. PROMIS: Patient-Reported Outcome Measurement Information System; VR: virtual reality.



Discussion

This report describes the 6-month durability of treatment effects for a randomized placebo-controlled trial of an 8-week self-administered skills-based VR program (EaseVRx) compared with a sham VR program in adults with CLBP. Intention-to-treat analysis performed on data collected 6 months after treatment revealed some regression to the mean with continued superiority of therapeutic EaseVRx over sham VR for reductions in pain intensity and pain-related interference (activity, stress, and sleep). Six-month posttreatment results exceeded thresholds for clinical meaningfulness, with effect sizes ranging from 0.34 to 0.54. Between-group differences for physical function and sleep disturbance at 6 months were statistically significant but not clinically meaningful. Combined, the results support the 6-month analgesic efficacy of a fully automated, 8-week, home-based VR program for CLBP. Recent meta-analyses of VR noted a lack of high-quality efficacy studies for chronic pain [21], except for those involving physical rehabilitation programs [22]. To our knowledge, our investigations on the extended efficacy of VR are the first involving home-based pain management without physical rehabilitation.

Findings from this study further support the efficacy of home-based VR treatment and may inform clinician and patient expectations, reimbursement models, and prescription pathways for CLBP. Critics have questioned whether participant education or socioeconomic factors might predict user engagement. Accordingly, we examined whether participant education level (high school level or less vs at least some college education) or household annual income (above vs below the US median) as a composite metric of SES would impact treatment engagement. While our examination of the impact of SES on user engagement is preliminary and may be subject to selection bias, we found equivalent engagement between lower and higher SES individuals with EaseVRx. These data potentially refute a perception that a high-tech digital treatment, such as VR, may be infeasible in lower SES individuals, and suggest that digital therapeutics, like EaseVRx, represent an opportunity to reach CLBP patients in historically underserved areas. These data also align with our published EaseVRx usability ratings, in which this study sample indicated that the device was as easy to use as an iPhone [4].

Key strengths of this study include (1) randomized placebo-controlled design; (2) intention-to-treat analyses; (3) correction for multiplicity; (4) longitudinal design and data collection to 6 months posttreatment; and (5) participant blinding to treatment group.

Our findings should be placed in the context of several limitations. First, the study sample had low levels of depressive symptoms and was specific to CLBP. The sample was also mainly female and white, and had some college education, thus limiting the generalizability to the broader population. The study relied only on participant-reported data and no objective data on medical or mental health conditions or receipt of additional pain treatments during the study period. The 26% attrition rate at 6 months was similar between treatment groups, and its effects were mitigated by the intention-to-treat analytic approach. Finally, at the end of the study roughly 73% of the sample correctly guessed their treatment group assignment, suggesting that, despite extensive efforts to maintain face validity of sham VR and following published guidance [8], the actual blinding failed. Despite this, we previously reported equivalent treatment engagement between both groups and symptom benefits gained by sham VR participants [7], albeit sham VR was substantially less efficacious than therapeutic VR. Nevertheless, equivalent engagement in sham VR suggests acceptable control in terms of time and attention, exposure to treatment, device use, survey completion, and participant compensation.

XSL•FO
RenderX

Placebo effects are well known in clinical studies [23]. The clinically meaningful efficacy of open-label placebo supports our finding of sham VR benefits even when treatment group assignment is correctly guessed [24]. The superiority and durability of the therapeutic response to therapeutic immersive VR is even more intriguing in the context of placebo interference with analgesic outcomes.

In-progress research includes an active national pragmatic effectiveness study designed to ascertain the long-term treatment effects of therapeutic VR in patients with CLBP who are highly diverse in race, ethnicity, education level, and symptom profiles. Future research should extend efficacy investigations for home-based VR to other pain conditions and diagnoses, as well as examine mechanisms of treatment effects in real-world patient populations. Finally, while we previously reported very high treatment engagement rates for both study groups, future research may investigate the characteristics and needs of individuals who exhibit lower engagement rates, nonresponsiveness, or higher rates of regression to the mean over time, and develop strategies to optimize outcomes for these subpopulations.

The 6-month durability of clinically meaningful reductions in pain intensity and pain-related interference suggests that this effective digital therapeutic approach may transcend many current barriers and improve patient access to effective nonpharmacologic pain care for CLBP.

Acknowledgments

AppliedVR, Inc supported this study. We thank Tracie Kim for her graphical support with the figures in this article and Dr Liesl Oldstone for numerous insightful comments on this work.

Data Availability

A data dictionary and deidentified participant data will be made available after publication and upon approved request of a detailed meta-analytic study proposal. Requests should be made to the corresponding author along with a study proposal and a signed data access agreement.

Authors' Contributions

LG was involved in study design, participant management, and data interpretation. BB was involved in study design. PK was involved in data analysis, interpretation, and manuscript preparation. IM was involved in participant management. JS was involved in project management. VS was involved in medication analysis. RL was involved in study design. CC was involved in participant management. RM was involved in data presentation and manuscript preparation. TM was involved in study design, data analysis, and manuscript preparation. BD was involved in study design, data interpretation, and manuscript preparation. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

TM and RM are employees of AppliedVR, Inc. JS is the president of AppliedVR, Inc. BB, PK, LG, IM, and VS are consultants for AppliedVR, Inc. CC is a contractor for AppliedVR, Inc. BD is the chief science advisor for AppliedVR, Inc. BD has authored or coauthored 5 pain treatment books for patients and clinicians, and receives royalties for 4 books. BD is the principal investigator for pain research grants and awards from the National Institutes of Health (NIH) and the Patient-Centered Research Outcomes Research Institute (none specific to the current work). BD is a co-investigator on 2 NIH research grants investigating virtual reality analgesia; neither of these grants is specific to the current work. BD serves on the Board of Directors for the American Academy of Pain Medicine and is on the Board of Directors for the Institute for Brain Potential. BD is a scientific member of the NIH Interagency Pain Research Coordinating Committee, the Centers for Disease Control and Prevention (CDC) Opioid Workgroup (2020-2021), and the Pain Advisory Group of the American Psychological Association.

References

1. Spiegel B, Fuller G, Lopez M, Dupuy T, Noah B, Howard A, et al. Virtual reality for management of pain in hospitalized patients: A randomized comparative effectiveness trial. PLoS One 2019;14(8):e0219115 [FREE Full text] [doi: 10.1371/journal.pone.0219115] [Medline: 31412029]
2. Atzori B, Hoffman HG, Vagnoli L, Patterson DR, Alhalabi W, Messeri A, et al. Virtual Reality Analgesia During Venipuncture in Pediatric Patients With Onco-Hematological Diseases. Front Psychol 2018 Dec 20;9:2508 [FREE Full text] [doi: 10.3389/fpsyg.2018.02508] [Medline: 30618938]
3. Darnall BD, Krishnamurthy P, Tsuei J, Minor JD. Self-Administered Skills-Based Virtual Reality Intervention for Chronic Pain: Randomized Controlled Pilot Study. JMIR Form Res 2020 Jul 07;4(7):e17293 [FREE Full text] [doi: 10.2196/17293] [Medline: 32374272]
4. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality Program for Chronic Low Back Pain: Double-Blind, Randomized,

XSL•FO
RenderX

Placebo-Controlled Trial Conducted During COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
5. Brea-Gómez B, Torres-Sánchez I, Ortiz-Rubio A, Calvache-Mateo A, Cabrera-Martos I, López-López L, et al. Virtual Reality in the Treatment of Adults with Chronic Low Back Pain: A Systematic Review and Meta-Analysis of Randomized Clinical Trials. Int J Environ Res Public Health 2021 Nov 11;18(22):11806 [FREE Full text] [doi: 10.3390/ijerph182211806] [Medline: 34831562]
6. Garcia LM, Darnall BD, Krishnamurthy P, Mackey IG, Sackman J, Louis RG, et al. Correction: Self-Administered Behavioral Skills-Based At-Home Virtual Reality Therapy for Chronic Low Back Pain: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2021 Feb 12;10(2):e27652 [FREE Full text] [doi: 10.2196/27652] [Medline: 33577465]
7. Garcia LM, Birckhead BJ, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, et al. Three-Month Follow-Up Results of a Double-Blind, Randomized Placebo-Controlled Trial of 8-Week Self-Administered At-Home Behavioral Skills-Based Virtual Reality (VR) for Chronic Low Back Pain. J Pain 2021 Dec 11 [FREE Full text] [doi: 10.1016/j.jpain.2021.12.002] [Medline: 34902548]
8. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
9. Sullivan M, Bishop S, Pivik J. The Pain Catastrophizing Scale: Development and validation. Psychological Assessment 1995;7(4):524-532. [doi: 10.1037/1040-3590.7.4.524]
10. Nicholas MK, McGuire BE, Asghari A. A 2-item short form of the Pain Self-efficacy Questionnaire: development and psychometric evaluation of PSEQ-2. J Pain 2015 Feb;16(2):153-163. [doi: 10.1016/j.jpain.2014.11.002] [Medline: 25463701]
11. Fish RA, McGuire B, Hogan M, Morrison TG, Stewart I. Validation of the chronic pain acceptance questionnaire (CPAQ) in an Internet sample and development and preliminary validation of the CPAQ-8. Pain 2010 Jun;149(3):435-443. [doi: 10.1016/j.pain.2009.12.016] [Medline: 20188472]
12. Nassif TH, Hull A, Holliday SB, Sullivan P, Sandbrink F. Concurrent Validity of the Defense and Veterans Pain Rating Scale in VA Outpatients. Pain Med 2015 Nov 01;16(11):2152-2161. [doi: 10.1111/pme.12866] [Medline: 26257151]
13. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther 2011;13(5):R147 [FREE Full text] [doi: 10.1186/ar3461] [Medline: 21914216]
14. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™ sleep disturbance and Sleep-Related Impairment item banks. Behav Sleep Med 2011 Dec 28;10(1):6-24 [FREE Full text] [doi: 10.1080/15402002.2012.636266] [Medline: 22250775]
15. Gershon R, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. J Appl Meas 2010;11(3):304-314 [FREE Full text] [Medline: 20847477]
16. Fischer HF, Rose M. Scoring Depression on a Common Metric: A Comparison of EAP Estimation, Plausible Value Imputation, and Full Bayesian IRT Modeling. Multivariate Behav Res 2019 Sep 20;54(1):85-99. [doi: 10.1080/00273171.2018.1491381] [Medline: 30235003]
17. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
18. Dworkin R, Turk D, McDermott M, Peirce-Sandner S, Burke L, Cowan P, et al. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009 Dec;146(3):238-244. [doi: 10.1016/j.pain.2009.08.019] [Medline: 19836888]
19. d'Errico A, Ricceri F, Stringhini S, Carmeli C, Kivimaki M, Bartley M, LIFEPATH Consortium. Socioeconomic indicators in epidemiologic research: A practical example from the LIFEPATH study. PLoS One 2017 May 30;12(5):e0178071 [FREE Full text] [doi: 10.1371/journal.pone.0178071] [Medline: 28557991]
20. Shrider E, Kollar M, Chen F, Semega J. Income and Poverty in the United States: 2020. The United States Census Bureau. 2021. URL: https://www.census.gov/library/publications/2021/demo/p60-273.html [accessed 2022-05-09]
21. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. JPR 2019 Jul;12:2053-2085. [doi: 10.2147/jpr.s200498]
22. Ahern MM, Dean LV, Stoddard CC, Agrawal A, Kim K, Cook CE, et al. The Effectiveness of Virtual Reality in Patients With Spinal Pain: A Systematic Review and Meta-Analysis. Pain Pract 2020 Jul 21;20(6):656-675. [doi: 10.1111/papr.12885] [Medline: 32196892]
23. Colloca L. The Placebo Effect in Pain Therapies. Annu Rev Pharmacol Toxicol 2019 Jan 06;59(1):191-211 [FREE Full text] [doi: 10.1146/annurev-pharmtox-010818-021542] [Medline: 30216744]
24. Carvalho C, Caetano JM, Cunha L, Rebouta P, Kaptchuk TJ, Kirsch I. Open-label placebo treatment in chronic low back pain: a randomized controlled trial. Pain 2016 Dec;157(12):2766-2772 [FREE Full text] [doi: 10.1097/j.pain.0000000000000700] [Medline: 27755279]

XSL•FO
RenderX

Abbreviations

CLBP: chronic low back pain
DVPRS: Defense and Veterans Pain Rating Scale
PROMIS: Patient-Reported Outcome Measurement Information System
SES: socioeconomic status
VR: virtual reality

Edited by G Eysenbach; submitted 22.02.22; peer-reviewed by P Gazerani; comments to author 15.03.22; revised version received 22.03.22; accepted 26.04.22; published 25.05.22.

Please cite as:
Garcia L, Birckhead B, Krishnamurthy P, Mackey I, Sackman J, Salmasi V, Louis R, Castro C, Maddox R, Maddox T, Darnall BD
Durability of the Treatment Effects of an 8-Week Self-administered Home-Based Virtual Reality Program for Chronic Low Back Pain: 6-Month Follow-up Study of a Randomized Clinical Trial
J Med Internet Res 2022;24(5):e37480
URL: https://www.jmir.org/2022/5/e37480
doi:10.2196/37480
PMID:35612905

©Laura Garcia, Brandon Birckhead, Parthasarathy Krishnamurthy, Ian Mackey, Josh Sackman, Vafi Salmasi, Robert Louis, Carina Castro, Roselani Maddox, Todd Maddox, Beth D Darnall. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 25.05.2022. 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.

XSL•FO
RenderX

Letter to the Editor

Application of Nonfungible Tokens to Health Care. Comment on "Blockchain Technology Projects to Provide Telemedical Services: Systematic Review"

John Gambril[1,2], MD; Carter Boyd[3], MD, MBA; Jamal Egbaria[4], MD

[1]Department of Internal Medicine, Ohio State University Wexner Medical Center, Columbus, OH, United States

[2]Department of Pediatrics, Nationwide Children's Hospital, Columbus, OH, United States

[3]Hansjörg Wyss Department of Plastic Surgery, New York University Langone Health, New York, NY, United States

[4]Department of Anesthesiology and Perioperative Medicine, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Corresponding Author:
John Gambril, MD
Department of Internal Medicine
Ohio State University Wexner Medical Center
410 W 10th Ave
Columbus, OH, 43210
United States
Phone: 1 614 293 8000
Email: alangambril@gmail.com

Related Article:

Comment on: http://www.jmir.org/2021/8/e17475/

(J Med Internet Res 2022;24(5):e34276) doi:10.2196/34276

KEYWORDS

telemedicine; distributed ledger; health information exchange; blockchain; cryptocurrency; nonfungible token; non-fungible token; medical education; internet; finance

In their systematic review, Koshechkin and colleagues [1] thoughtfully explored and summarized the potential applications of blockchain technology to solve unique challenges in health care. Blockchain technology is a highly secure but transparent method of tracing digital transactions of assets or information through a decentralized, immutable ledger [1]. Thanks to the expansion of cryptocurrency—digital currency traded through blockchain networks—blockchain has become a household term. Blockchain technology has been applied to numerous industries, and many possibilities exist for application within health care [2].

In their review, Koshechkin and colleagues [1] identified 18 studies addressing blockchain solutions for various health care challenges. These included medical data access, medical services processing, diagnostic support, payment transactions, and fundraising, among others. This review offers an excellent synopsis of ongoing blockchain projects in health care accompanied by discussion of future directions.

One novel entity that utilizes blockchain technology not identified in this paper is the nonfungible token (NFT). NFTs are similar to cryptocurrency, such as Bitcoin, in that they are digital tokens existing within a blockchain that can be bought and sold. NFTs, however, are distinct tokens. While 1 Bitcoin is always equal to another Bitcoin, each NFT is singular and unique (ie, nonfungible) [3]. Interestingly, although NFTs are digital, they can, and often do, represent physical items.

Regarding health care, NFTs have been suggested as a means to streamline and simplify blood [4] and stem cell [5] product supply chains. While the full discussion of NFT potential in health care warrants a lengthier article, we will touch on the simplest application: capital.

NFTs represent an untapped method of fundraising and revenue generation. They allow the monetization of unique items and content, digital or physical, new or old. NFTs offer the added benefits of transaction tracing, verifiable authenticity, and shared ownership of physical objects (analogous to stockholders of a company). One can imagine NFT auctions of historic medical equipment or journal articles, recorded lectures from renowned experts, and even naming rights to a lecture series. Much like baseball cards or stamps, medicine could find its own niche in collecting NFTs.

XSL•FO
RenderX

NFTs also open an avenue for monetization of original creations such as reflective art and literature or unique medical education content. Medical education resources are constantly evolving, and many are free to learners (tweetorials, podcasts, YouTube channels, etc). If collectors agree to maintain open access after purchase, leveraging this content as NFTs could represent a new market to raise funds for scholarships, research, advocacy, or public health projects. Many will argue that buyers are unlikely to permit open access to their NFT. However, as already seen in other NFT markets, because NFTs are unique and traceable digital tokens, ownership can still be boasted despite public access [3].

As illustrated by Koshechkin and colleagues' [1] review, blockchain technology is an exciting entity offering much potential for the advancement of health care. NFTs are yet another example of blockchain technology with innumerable possibilities and potential in health care.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Blockchain Technology Projects to Provide Telemedical Services: Systematic Review" was invited to provide an Author Reply but did not provide a response to this Letter to the Editor prior to publication.

References

1. Koshechkin K, Lebedev G, Radzievsky G, Seepold R, Martinez NM. Blockchain technology projects to provide telemedical services: Systematic review. J Med Internet Res 2021 Aug 18;23(8):e17475 [FREE Full text] [doi: 10.2196/17475] [Medline: 34407924]
2. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
3. Conti R, Schmidt J. What you need to know about non-fungible tokens (NFTs). Forbes Advisor. 2021 May 14. URL: https://www.forbes.com/advisor/investing/nft-non-fungible-token/ [accessed 2021-10-10]
4. McFarlane C. Tokenized blood? How NFTs are transforming healthcare. Forbes. 2021 Jun 21. URL: https://www.forbes.com/sites/chrissamcfarlane/2021/06/02/will-nfts-save-healthcare/?sh=1efae86ceae3 [accessed 2021-10-10]
5. Booth GS, Gehrie EA. Non-fungible tokens: Stem cell transplantation in the blockchain. Transfus Apher Sci 2021 Oct;60(5):103196. [doi: 10.1016/j.transci.2021.103196] [Medline: 34148765]

Abbreviations

NFT: nonfungible token

Edited by T Leung; submitted 14.10.21; this is a non–peer-reviewed article;accepted 05.05.22; published 30.05.22.

Please cite as:
Gambril J, Boyd C, Egbaria J
Application of Nonfungible Tokens to Health Care. Comment on "Blockchain Technology Projects to Provide Telemedical Services: Systematic Review"
J Med Internet Res 2022;24(5):e34276
URL: https://www.jmir.org/2022/5/e34276
doi:10.2196/34276
PMID:35635749

©John Gambril, Carter Boyd, Jamal Egbaria. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 30.05.2022. 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.

XSL•FO
RenderX

Letter to the Editor

Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on "Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial"

Chih-Wei Chen[1], MPhil; James Cheng-Chung Wei[1,2,3], MD, PhD

[1]Institute of Medicine, Chung Shan Medical University, Taichung, Taiwan

[2]Department of Allergy, Immunology & Rheumatology, Chung Shan Medical University Hospital, Taichung, Taiwan

[3]Graduate Institute of Integrated Medicine, China Medical University, Taichung, Taiwan

Corresponding Author:
James Cheng-Chung Wei, MD, PhD
Institute of Medicine
Chung Shan Medical University
No 110, Sec 1
Jianguo N Rd, South District
Taichung, 40201
Taiwan
Phone: 886 4 24739595
Fax: 886 4 24637389
Email: jccwei@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2021/4/e23446

Comment in: https://www.jmir.org/2022/5/e37413/

(J Med Internet Res 2022;24(5):e34437) doi:10.2196/34437

KEYWORDS

telemedicine; oxygen therapy; noninvasive positive airway pressure; BiPAP; osteopathic medicine; physical therapy; SARS-CoV-2; COVID-19; teletherapy; telemanagement

We read with great interest the research article by Adly et al [1] regarding the nonpharmacological respiratory treatment methods for home-isolated patients with COVID-19, involving the use of a newly developed telemanagement health care system. We appreciate the authors' valuable contribution to providing a deeper understanding of the effectiveness of home-based oxygen therapy with bilevel positive airway pressure (BiPAP) and that of osteopathic manipulative respiratory and physical therapy techniques in impeding exacerbation of early-stage COVID-19 pneumonia. However, the study should be further discussed in terms of research design, patient recruitment technologies, and the influence of the telemanagement system on health care development.

To begin with, the researchers randomly recruited 60 patients for their study through social media by using a snowball subject recruitment technique. However, the health conditions of these patients (eg, age and underlying co-morbidities), which significantly affect their outcomes [2], were not considered in the analysis. Moreover, the eligibility of recruited patients has not been well addressed in their publication. The sample size and power calculation should also be presented in a clinical trial protocol.

Moreover, the authors used the snowball subject recruitment technique through social media to recruit patients, which was completely random and could not ensure that all patients met the eligibility criteria. However, successful patient recruitment requires a rational clinical design, efficient patient identification and randomization, which can be fulfilled by various information technologies such as data mining, artificial intelligence, and automated alerts [3]. Furthermore, the echo chamber effect on social media could also result in the limitation of patient recruitment [4]. Hence, we suggest that the authors use suitable

XSL•FO
RenderX

technologies based on a comprehensive database for patient recruitment.

Furthermore, the national health care database has been increasingly developed and used as a comprehensive database for clinical trials [5]. The employment of the telemanagement system in home-based treatment for patients with COVID-19 allows for data collection through the system and uses fewer human resources. The integration of the national health care database and the telemanagement system could thus allow health care workers to provide long-distance health care not only in the context of COVID-19 treatment but also in health care in the future society. Hence, the authors are suggested to discuss the influence of the implementation of telemanagement together with the national health care database on health care development.

Above all, this study contributes to providing a deeper understanding of the effectiveness of home-based oxygen therapy with BiPAP, and that of the osteopathic manipulative respiratory and physical therapy techniques in impeding exacerbation of early-stage COVID-19 pneumonia. Further studies would likely enhance the research design, implement patient recruitment by more suitable technologies, and discuss the influence of the telemanagement system on health care development.

Conflicts of Interest

None declared.

References

1. Adly AS, Adly MS, Adly AS. Telemanagement of home-isolated COVID-19 patients using oxygen therapy with noninvasive positive pressure ventilation and physical therapy techniques: randomized clinical trial. J Med Internet Res 2021;23(4):e23446. [doi: 10.2196/23446]
2. Semenzato L, Botton J, Drouin J, Cuenot F, Dray-Spira R, Weill A, et al. Chronic diseases, health conditions and risk of COVID-19-related hospitalization and in-hospital mortality during the first wave of the epidemic in France: a cohort study of 66 million people. Lancet Reg Health Eur 2021;8(100158). [doi: doi: 10.1016/j.lanepe.2021.100158]
3. von Itzstein MS, Hullings M, Mayo H, Beg MS, Williams EL, Gerber DE. Application of information technology to clinical trial evaluation and enrollment: a review. JAMA Oncol 2021;7(10):1559-1566. [doi: 10.1001/jamaoncol.2021.1165] [Medline: 34236403]
4. Cinelli M, Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci USA 2021;118(9):e2023301118. [doi: 10.1073/pnas.2023301118]
5. Yong SB, Huang JY, Wei JC. Adult outcome of juvenile idiopathic arthritis: A nationwide population-based retrospective cohort study in Taiwan. Int J Rheum Dis 2019;22(7):1283-1288. [doi: 10.1111/1756-185X.13527]

Abbreviations

BiPAP: bilevel positive airway pressure

Edited by T Leung; submitted 22.10.21; this is a non–peer-reviewed article;accepted 26.04.22; published 02.05.22.

Please cite as:
Chen CW, Wei JCC
Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on "Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial"
J Med Internet Res 2022;24(5):e34437
URL: https://www.jmir.org/2022/5/e34437
doi:10.2196/34437
PMID:35476637

©Chih-Wei Chen, James Cheng-Chung Wei. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.05.2022. 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.

XSL•FO
RenderX

Letter to the Editor

Authors Reply to: Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on "Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial"

Aya Sedky Adly[1,2], MSc, PhD; Mahmoud Sedky Adly[3,4], MSc, DDS, HDD; Afnan Sedky Adly[5,6], MSc

[1]Faculty of Engineering and Technology, Badr University in Cairo (BUC), Cairo, Egypt

[2]Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt

[3]Royal College of Surgeons of Edinburgh, Scotland, United Kingdom

[4]Faculty of Oral and Dental Medicine, Cairo University, Cairo, Egypt

[5]Faculty of Physical Therapy, Cardiovascular-Respiratory Disorders and Geriatrics, Laser Applications in Physical Medicine, Cairo University, Cairo, Egypt

[6]Faculty of Physical Therapy, Internal Medicine, Beni-Suef University, Cairo, Egypt

Corresponding Author:
Aya Sedky Adly, MSc, PhD
Faculty of Engineering and Technology, Badr University in Cairo (BUC)
Cairo Suez Road
Badr City
Cairo, 11511
Egypt
Phone: 20 1145559778
Email: aya.sedky@gmail.com

Related Articles:

Comment on: https://www.jmir.org/2022/5/e34437

Comment on: https://www.jmir.org/2021/4/e23446

(J Med Internet Res 2022;24(5):e37413) doi:10.2196/37413

KEYWORDS

telemedicine; oxygen therapy; noninvasive positive airway pressure; BiPAP; osteopathic medicine; physical therapy; SARS-CoV-2; COVID-19; teletherapy; telemanagement

We would like to thank the authors for their interest in reading our published paper [1,2], and we would like to clarify the following points.

First, the study design and patient recruitment technique have been illustrated in sufficient detail in the methods section. Although there is a part that discusses the influence of the telemanagement system on health care development, the aim of this study was to compare two nonpharmacological respiratory treatment methods for home-isolated patients with COVID-19 by using a newly developed telemanagement health care system. The influence of the telemanagement system on health care development has been further discussed in many other articles [3-6], and its discussion was therefore considered outside of the scope of our paper.

Second, concerning the health conditions of the study patients, we disagree with the comment because these variables were considered in both this study and its analysis. As noted in the inclusion criteria, only the American Society of Anesthesiologists Class-I patients were deemed eligible for participation in the study before onset of COVID-19. We have also stated in the results section that the patients' age ranged between 21 and 40 years.

Third, the eligibility of the recruited patients, inclusion criteria, and exclusion criteria have already been well addressed in the methods section by specifically defining the selection criteria to avoid any unwanted effect of confounding factors.

XSL•FO
RenderX

Fourth, both the sample size and power calculations have already been presented in the statistical analysis section.

Fifth, we also disagree with the comment concerning the snowball subject recruitment technique, since it is considered an efficient sampling technique that is used to recruit participants who are hard to be located.

Sixth, we have reviewed the mentioned reference about the echo chamber effect on social media and found that it discusses public opinion formation and not patient recruitment. Thus, for this suggestion, we would expect to be provided with evidence on the suitability of a comprehensive database for patient recruitment. We would also want to clarify that all the selected patients met the study's eligibility criteria, and in the methods section, we have explained how we ensured this.

Seventh, studying the influence of the implementation of telemanagement together with the national health care database on health care development was not the aim of this study.

Eighth, we believe that the research design was thoroughly explained with sufficient detail that can be easily understood by any health care worker. However, for a suggestion on design enhancement, we would expect to be more specific on what points the design should be enhanced, along with an acceptable reason.

Moreover, for the suggestion to implement patient recruitment using more suitable techniques, we would expect to be provided with evidence that the patient recruitment technique used in our study is inferior to other techniques or that it has a negative effect on the study outcomes, which is not the case.

In addition, the influence of the telemanagement system on health care development has been addressed by many studies and is beyond the scope of this publication.

Finally, we appreciate this chance to reply to this letter and would like to advise researchers to consider recruiting larger study samples before adopting our techniques.

Conflicts of Interest

None declared.

References

1. Adly AS, Adly MS, Adly AS. Telemanagement of home-isolated COVID-19 patients using oxygen therapy with noninvasive positive pressure ventilation and physical therapy techniques: randomized clinical trial. J Med Internet Res 2021;23(4):e23446. [doi: 10.2196/23446]
2. Chen CW, Wei JCC. Research using social media to recruit research participants should proceed with caution. Comment on "Telemanagement of home-isolated COVID-19 patients using oxygen therapy with noninvasive positive pressure ventilation and physical therapy techniques: randomized clinical trial". J Med Internet Res 2022;24(5):e34437. [doi: 10.2196/34437]
3. Adly AS, Adly AS, Adly MS. Effects of laser acupuncture tele-therapy for rheumatoid arthritis elderly patients. Lasers Med Sci 2022;37:499-504. [doi: 10.1007/s10103-021-03287-0]
4. Adly MS, Adly AS, Alreshidi SF, Alotaibi AM, Alreshidi MF, Adly AS. Can paracetamol lower stress and anxiety by blunting emotions during and after computer guided dental implant surgeries? Findings from a randomized crossover clinical trial. Journal of dental sciences 2021;16:137-144. [doi: 10.1016/j.jds.2020.05.026]
5. Adly AS, Adly AS, Adly MS, Ali MF. A novel approach utilizing laser acupuncture teletherapy for management of elderly-onset rheumatoid arthritis: A randomized clinical trial. J Telemed Telecare 2021;27:298-306. [doi: 10.1177/1357633X211009861]
6. Adly AS, Adly AS, Adly MS. Healthcare technology trade-offs for IoT ecosystems from a developing country perspective: case of Egypt. Healthcare Paradigms in the Internet of Things Ecosystem: Academic Press 2021:1-13. [doi: 10.1016/B978-0-12-819664-9.00014-4]

Edited by T Leung; submitted 19.02.22; this is a non–peer-reviewed article;accepted 26.04.22; published 02.05.22.

Please cite as:
Adly AS, Adly MS, Adly AS
Authors Reply to: Research Using Social Media to Recruit Research Participants Should Proceed With Caution. Comment on "Telemanagement of Home-Isolated COVID-19 Patients Using Oxygen Therapy With Noninvasive Positive Pressure Ventilation and Physical Therapy Techniques: Randomized Clinical Trial"
J Med Internet Res 2022;24(5):e37413
URL: https://www.jmir.org/2022/5/e37413
doi:10.2196/37413
PMID:35476751

XSL•FO
RenderX

©Aya Sedky Adly, Mahmoud Sedky Adly, Afnan Sedky Adly. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 02.05.2022. 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.

XSL•FO
RenderX

Letter to the Editor

Learning More About the Effects of Gamification on Physical Activity. Comment on "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials"

Cheng-Hsien Hung[1], BSc; Yung-Shuo Kao[2], MD

[1]Department of Pharmacy, Chang Bing Show Chwan Memorial Hospital, Changhua County, Taiwan

[2]Department of Radiation Oncology, China Medical University Hospital, Taichung, Taiwan

Corresponding Author:
Cheng-Hsien Hung, BSc
Department of Pharmacy
Chang Bing Show Chwan Memorial Hospital
No 6, Lugong Rd
Lukang Township
Changhua County, 50544
Taiwan
Phone: 886 4 7813888 ext 71255
Fax: 886 4 7812966
Email: chenghsien823@gmail.com

Related Articles:

Comment on: http://www.jmir.org/2022/1/e26779/

Comment in: http://mhealth.jmir.org/2022/5/e38212/

(J Med Internet Res 2022;24(5):e36396) doi:10.2196/36396

KEYWORDS

behavior change; eHealth; gamification; health behavior; intervention; meta-analysis; mobile phone; physical activity; systematic review; elderly; old adults

We read with great interest the article, "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials" by Mazeas et al [1]. The authors conducted a systematic review and meta-analysis to evaluate the effectiveness of gamification on physical activity (PA). This meta-analysis confirms that gamified interventions are promising for promoting PA in various populations. Although the authors acknowledge the limitations of their study, we wish to highlight several methodological issues and provide our perspective.

First, Paul et al [2] was not a randomized controlled trial, but a nonrandomized clinical controlled trial. As mentioned in both the methods and limitations of this study, the authors used nonrandomized allocation. According to the Cochrane Handbook for Systematic Reviews of Interventions [3], "Predefined, unambiguous eligibility criteria are a fundamental prerequisite for a systematic review." Authors should select literature strictly according to inclusion and exclusion criteria, especially for the quantitative analysis. Mixing research with different study designs may significantly affect the results and the level of evidence.

Second, the total number of hours of gamification performed can make a significant difference. Although the authors have conducted a subgroup analysis of the duration of gamification (short- and long-term interventions), each type of gamification is different, and we do not know the number of minutes of gamification performed per day. The total number of hours of gamification performed can vary greatly across the different studies. We believe this could be a potential source of heterogeneity.

Third, this meta-analysis may not apply to older adults. The mean age of participants in the selected studies was 35.7 years, and most of the studies were conducted on participants aged <65 years. The American College of Sports Medicine suggests that the population most in need of exercise may be older adults [4]. In a previous meta-analysis, the lack of PA was associated with all-cause mortality and cardiovascular mortality, fractures,

XSL•FO
RenderX

and falls among older populations [5]. However, this age group may have difficulty with device operation and gamification rules. Age may affect the gamification experience, leading to limitations in the application of evidence.

In conclusion, we believe that clarification of the above points can strengthen the interpretation of the study results. The authors have analyzed an important issue. A better understanding of the effects of gamification mechanisms on PA is critical for clinicians.

Conflicts of Interest

None declared.

References

1. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
2. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
3. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). 2021. URL: https://training.cochrane.org/handbook [accessed 2021-01-12]
4. American College of Sports Medicine, Chodzko-Zajko WJ, Proctor DN, Fiatarone Singh MA, Minson CT, Nigg CR, et al. American College of Sports Medicine position stand. Exercise and physical activity for older adults. Med Sci Sports Exerc 2009 Jul;41(7):1510-1530. [doi: 10.1249/MSS.0b013e3181a0c95c] [Medline: 19516148]
5. Cunningham C, O' Sullivan R, Caserotti P, Tully MA. Consequences of physical inactivity in older adults: A systematic review of reviews and meta-analyses. Scand J Med Sci Sports 2020 May;30(5):816-827. [doi: 10.1111/sms.13616] [Medline: 32020713]

Abbreviations

PA: physical activity

Edited by T Leung; submitted 13.01.22; this is a non–peer-reviewed article;accepted 26.04.22; published 03.05.22.

Please cite as:
Hung CH, Kao YS
Learning More About the Effects of Gamification on Physical Activity. Comment on "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials"
J Med Internet Res 2022;24(5):e36396
URL: https://www.jmir.org/2022/5/e36396
doi:10.2196/36396
PMID:35503654

©Cheng-Hsien Hung, Yung-Shuo Kao. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2022. 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.

XSL•FO
RenderX

Letter to the Editor

Authors' Reply to: Learning More About the Effects of Gamification on Physical Activity. Comment on "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials"

Alexandre Mazeas[1,2,3], MSc; Martine Duclos[2,4], PhD; Bruno Pereira[5], PhD; Aïna Chalabaev[1], PhD

[1]Univ Grenoble Alpes, SENS, 38000 Grenoble, France

[2]National Research Institute for Agriculture, Food and Environment (INRAE), Clermont-Ferrand, France

[3]Kiplin, Nantes, France

[4]Department of Sport Medicine and Functional Exploration, University Hospital Clermont-Ferrand, Hospital G Montpied, Clermont-Ferrand, France

[5]Department of Biostatistics Unit, Clermont-Ferrand University Hospital, Clermont-Ferrand, France

Corresponding Author:
Alexandre Mazeas, MSc
Univ Grenoble Alpes
SENS
1741 rue de la Piscine
38000 Grenoble
France
Phone: 33 476635081
Email: alexandre.mazeas@univ-grenoble-alpes.fr

Related Articles:

Comment on: http://www.jmir.org/2022/5/e36396/

Comment on: http://www.jmir.org/2022/1/e26779/

(J Med Internet Res 2022;24(5):e38212) doi:10.2196/38212

KEYWORDS

behavior change; eHealth; gamification; health behavior; intervention; meta-analysis; mobile phone; physical activity; systematic review; elderly; old adults

We appreciated and read with attention Hung and Kao's [1] feedback on our recent systematic review and meta-analysis [2], which examined the effectiveness of gamified interventions on physical activity. These authors have pointed out 3 aspects that we will discuss in this letter.

First, they suggest that Paul et al [3] should not have been included in our review as this study is a nonrandomized clinical controlled trial. We agree that this study was nonrandomized. We have in fact mentioned this issue under the Risk of Bias subheading of our Results section: "Overall, 1 study [28] was rated as high risk for sequence generation because assignments were based on recruitment order," where reference 28 points to Paul et al [3]. This statement was also reported in Multimedia Appendix 2 and was taken into consideration in the summary of findings following the GRADE (grading of recommendations assessment, development, and evaluation) framework, where the quality of evidence for some meta-analyses was downgraded because of the risks of bias in the included studies. Thus, these limitations have been taken into account in our review. Moreover, we would like to emphasize that Paul et al's [3] study did not have a large heterogeneity contribution and effect size influence as highlighted by our leave-one-out analyses and Baujat plot available in Multimedia Appendix 1. As an example, when omitting this study from the final sample (ie, after sensitivity analyses), we obtained a Hedges *g* of 0.40 (95% CI 0.11-0.75).

Second, Hung and Kao [1] suggest that the total number of hours of gamification performed can have a significant influence and could explain heterogeneity. We cannot agree more on this point since we are convinced that engagement with digital behavior change interventions is necessary to enable an effective intervention. Gamification has often been assimilated into a self-fulfilling process permitting automatic engagement of participants into an eHealth service. However, this is not always the case, which can influence the effect of the intervention. Nevertheless, very few studies measured both engagement and

XSL•FO
RenderX

behavioral outcomes in the included studies, preventing us from examining the possible existence of a dose-response. Therefore, we would recommend that future trials should systematically combine measures of engagement in addition to other outcomes. Engagement with the gamified service can be objectively recorded using data from apps and websites (eg, number of logins, time spent per login, number of components accessed), measured via self-report questionnaires (eg, the DBCI Engagement Scale [4]), psychological measures of attention, and qualitative or observational methods.

Finally, Hung and Kao [1] also noted that the results of this meta-analysis may not apply to older adults. If through our meta-regression, the age of participants was not statistically significantly associated with the intervention effect, it is clear that our conclusions cannot be generalized to participants outside the age scope of our review (9-73 years). As they pointed out, few studies have evaluated the effect of gamified interventions on older adults. Future studies should focus on this specific population with specific characteristics.

Conflicts of Interest

None declared.

References

1. Hung CH, Kao YS. Learning More About the Effects of Gamification on Physical Activity. Comment on "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials". J Med Internet Res 2022 May;24(5):e36396 [FREE Full text] [doi: 10.2196/36396]
2. Mazeas A, Duclos M, Pereira B, Chalabaev A. Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022 Jan 04;24(1):e26779 [FREE Full text] [doi: 10.2196/26779] [Medline: 34982715]
3. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
4. Perski O, Blandford A, Garnett C, Crane D, West R, Michie S. A self-report measure of engagement with digital behavior change interventions (DBCIs): development and psychometric evaluation of the "DBCI Engagement Scale". Transl Behav Med 2020 Feb 03;10(1):267-277 [FREE Full text] [doi: 10.1093/tbm/ibz039] [Medline: 30927357]

Abbreviations

GRADE: grading of recommendations assessment, development, and evaluation

Edited by T Leung; submitted 24.03.22; this is a non–peer-reviewed article;accepted 26.04.22; published 03.05.22.

Please cite as:
Mazeas A, Duclos M, Pereira B, Chalabaev A
Authors' Reply to: Learning More About the Effects of Gamification on Physical Activity. Comment on "Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials"
J Med Internet Res 2022;24(5):e38212
URL: https://www.jmir.org/2022/5/e38212
doi:10.2196/38212
PMID:35503414

©Alexandre Mazeas, Martine Duclos, Bruno Pereira, Aïna Chalabaev. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 03.05.2022. 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.

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data From the Health Information National Trends Survey

Alexandra J Greenberg-Worisek[1], PhD, MPH; Shaheen Kurani[1], MSc; Lila J Finney Rutten[1], PhD, MPH; Kelly D Blake[2], SCD; Richard P Moser[2], PhD; Bradford W Hesse[2], PhD

[1]Mayo Clinic College of Medicine and Science, Rochester, MN, United States

[2]National Cancer Institute, National Institutes of Health, Rockville, MD, United States

Corresponding Author:
Alexandra J Greenberg-Worisek, PhD, MPH
Mayo Clinic College of Medicine and Science
200 First Street SW
Rochester, MN, 55906
United States
Phone: 1 507 538 7388
Email: worisek.alexandra@gmail.com

Related Article:

Correction of: https://www.jmir.org/2019/6/e13300/

(J Med Internet Res 2022;24(5):e39712) doi:10.2196/39712

In "Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data From the Health Information National Trends Survey" (J Med Internet Res 2019;21(6):e13300) the authors made the following updates.

The authors were notified of data errors in two of the Health Information National Trends Survey (HINTS) cycle datasets (HINTS 4, Cycle 3 and HINTS 4, Cycle 4); the errors were in the weights provided for use in the analysis of these data [1].

These weights primarily affected logistic regression analyses, reported in Table 2 of the originally published article. This previous version of Table 2 is in Multimedia Appendix 1. Following the HINTS error notice [1], the authors reran the logistic regression analyses. The corrected version of the article includes the following updated Table 2.

In rerunning this analysis, only one difference was found that resulted in changes in the conclusion. Namely, geography is significant, with those in rural settings having significantly lower odds of having internet access via a mobile phone compared to their urban counterparts (OR = 0.80; 95% CI: 0.65-0.98; *P*=.033). This finding was not reported as significant in the original analysis, due to the above-mentioned error in the HINTS data sets [1]. All other conclusions remain consistent with those reported in the original publication.

In the section "Internet Access via Cellular Network" in the *Results*, the first sentence in the second paragraph originally read as follows:

> *Most of the sociodemographic variables within our multivariable model were statistically significant after adjusting for survey year, save for geography (Table 2).*

It has been corrected as follows:

> *Most of the sociodemographic variables within our multivariable model were statistically significant after adjusting for survey year (Table 2).*

XSL•FO
RenderX

Table 2. Weighted multivariate logistic regression model of predictors of having internet access via mobile phone among those who reported having internet access. Data from the National Cancer Institute's Health Information National Trends Survey (HINTS) administrations between 2008 and 2017 (n=14,794).

Variable	Predictors of internet access via cell phone				
	Odds ratio (95% CI)	Beta coefficient	SE beta	Adjusted Wald F	*P* value
Sex				1.32	.252
Female	Ref[a]	Ref	Ref		
Male	1.08 (0.95, 1.22)	0.7	0.06		
Age				166.15	<.001
18-34	Ref	Ref	Ref		
35-49	0.43 (0.36-0.51)	–0.84	0.09		
50-64	0.20 (0.17-0.24)	–1.61	0.10		
65-74	0.08 (0.06-0.10)	–2.52	0.11		
>75	0.04 (0.03-0.05)	–3.24	0.16		
Race and ethnicity				4.07	.008
Non-Hispanic White	Ref	Ref	Ref		
Hispanic	1.25 (1.00-1.56)	0.23	0.11		
Non-Hispanic Black	1.39 (1.09-1.77)	0.33	0.12		
Non-Hispanic Other	0.83 (0.63-1.10)	–0.18	0.14		
Education				5.26	.002
Less than high school	Ref	Ref	Ref		
High school graduate	1.03 (0.65-1.64)	0.03	0.24		
Some college	1.42 (0.89-2.27)	0.35	0.24		
College graduate	1.47 (0.93-2.31)	0.38	0.23		
Income (US $)				14.06	<.001
<$20,000	Ref	Ref	Ref		
$20,000 to <$35,000	0.98 (0.73-1.30)	–0.02	0.15		
$35,000 to <$50,000	1.07 (0.83-1.39)	0.07	0.13		
$50,000 to <$75,000	1.33 (1.04-1.70)	0.28	0.13		
$75,000 +	1.92 (1.50-2.46)	0.65	0.12		
Geography				4.60	.033
Urban	Ref	Ref	Ref		
Rural	0.80 (0.65-0.98)	–0.23	0.11		
HINTS [b] Survey Year				126.77	<.001
HINTS 3 (2008)	Ref	Ref	Ref		
HINTS 4 Cycle 1 (2011)	17.86 (13.17-24.21)	2.88	0.15		
HINTS 4 Cycle 2 (2012)	21.59 (16.06-29.02)	3.07	0.15		
HINTS 4 Cycle 3 (2013)	29.45 (21.32-40.69)	3.38	0.16		
HINTS 4 Cycle 4 (2014)	30.45 (22.24-41.69)	3.42	0.16		
HINTS 5 Cycle 1 (2017)	51.31 (37.54-70.11)	3.94	0.16		

[a]Ref: reference group.

[b]HINTS: Health Information National Trends Survey.

In addition, the corresponding author's email address has been changed from *greenberg.alexandra@mayo.edu* to *worisek.alexandra@gmail.com*, as the author is no longer affiliated with Mayo Clinic College of Medicine and Science.

XSL•FO
RenderX

The correction will appear in the online version of the paper on the JMIR Publications website on May 26, 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.

Multimedia Appendix 1
Originally published Table 2.
[PNG File , 731 KB - jmir_v24i5e39712_app1.png]

Reference

1. HINTS Data Errors, Remediation, and Recommendations. Health Information National Trends Survey. URL: https://hints.cancer.gov/data/data-remediation.aspx [accessed 2022-05-19]

Abbreviations

HINTS: Health Information National Trends Survey

Submitted 19.05.22; this is a non–peer-reviewed article;accepted 19.05.22; published 26.05.22.

Please cite as:
Greenberg-Worisek AJ, Kurani S, Finney Rutten LJ, Blake KD, Moser RP, Hesse BW
Correction: Tracking Healthy People 2020 Internet, Broadband, and Mobile Device Access Goals: An Update Using Data From the Health Information National Trends Survey
J Med Internet Res 2022;24(5):e39712
URL: https://www.jmir.org/2022/5/e39712
doi:10.2196/39712
PMID:35617681

©Alexandra J Greenberg-Worisek, Shaheen Kurani, Lila J Finney Rutten, Kelly D Blake, Richard P Moser, Bradford W Hesse. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 26.05.2022. 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.

XSL•FO
RenderX

Original Paper

Monitoring User Opinions and Side Effects on COVID-19 Vaccines in the Twittersphere: Infodemiology Study of Tweets

Beatrice Portelli[1,2], MSc; Simone Scaboro[1], MSc; Roberto Tonino[1], BSc; Emmanuele Chersoni[3], PhD; Enrico Santus[4], PhD; Giuseppe Serra[1], PhD

[1]Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy

[2]Department of Biology, Università degli Studi di Napoli Federico II, Napoli, Italy

[3]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

[4]Decision Science and Advanced Analytics for Medical Affairs and Pharmacovigilance, Bayer Pharmaceuticals, Whippany, NJ, United States

Corresponding Author:
Beatrice Portelli, MSc
Department of Mathematics, Computer Science and Physics
University of Udine
206 via delle Scienze
Udine, 33100
Italy
Phone: 39 0432558400
Email: portelli.beatrice@spes.uniud.it

Abstract

Background: In the current phase of the COVID-19 pandemic, we are witnessing the most massive vaccine rollout in human history. Like any other drug, vaccines may cause unexpected side effects, which need to be investigated in a timely manner to minimize harm in the population. If not properly dealt with, side effects may also impact public trust in the vaccination campaigns carried out by national governments.

Objective: Monitoring social media for the early identification of side effects, and understanding the public opinion on the vaccines are of paramount importance to ensure a successful and harmless rollout. The objective of this study was to create a web portal to monitor the opinion of social media users on COVID-19 vaccines, which can offer a tool for journalists, scientists, and users alike to visualize how the general public is reacting to the vaccination campaign.

Methods: We developed a tool to analyze the public opinion on COVID-19 vaccines from Twitter, exploiting, among other techniques, a state-of-the-art system for the identification of adverse drug events on social media; natural language processing models for sentiment analysis; statistical tools; and open-source databases to visualize the trending hashtags, news articles, and their factuality. All modules of the system are displayed through an open web portal.

Results: A set of 650,000 tweets was collected and analyzed in an ongoing process that was initiated in December 2020. The results of the analysis are made public on a web portal (updated daily), together with the processing tools and data. The data provide insights on public opinion about the vaccines and its change over time. For example, users show a high tendency to only share news from reliable sources when discussing COVID-19 vaccines (98% of the shared URLs). The general sentiment of Twitter users toward the vaccines is negative/neutral; however, the system is able to record fluctuations in the attitude toward specific vaccines in correspondence with specific events (eg, news about new outbreaks). The data also show how news coverage had a high impact on the set of discussed topics. To further investigate this point, we performed a more in-depth analysis of the data regarding the AstraZeneca vaccine. We observed how media coverage of blood clot–related side effects suddenly shifted the topic of public discussions regarding both the AstraZeneca and other vaccines. This became particularly evident when visualizing the most frequently discussed symptoms for the vaccines and comparing them month by month.

Conclusions: We present a tool connected with a web portal to monitor and display some key aspects of the public's reaction to COVID-19 vaccines. The system also provides an overview of the opinions of the Twittersphere through graphic representations, offering a tool for the extraction of suspected adverse events from tweets with a deep learning model.

(J Med Internet Res 2022;24(5):e35115) doi:10.2196/35115

XSL•FO
RenderX

KEYWORDS

adverse drug events; COVID-19; digital pharmacovigilance; opinion mining; vaccines; social media; machine learning; deep learning; learning models; sentiment analysis; Twitter analysis; Twitter; web portal; public health

Introduction

Background

The COVID-19 pandemic has been at the heart of the discussions on all media outlets for almost 2 years. These debates touch upon very important and sensitive topics such as health, politics, work, school, and personal freedom to cite only a few. In a general effort to tackle the pandemic, many countries have engaged in the fastest and most massive vaccine rollout witnessed in human history: in less than 1 year, several vaccines have been created, tested, and distributed around the world, and many others are at the last phase of clinical trials and/or waiting for approval from regulatory agencies [1]. Despite the great efforts put into development, the rollout of vaccines has been slowed down in various countries [2] due to hesitancy and fake news poisoning social media debates. The vaccination rollout for the first strains of the virus has proceeded slower than initially planned, and experts agree that it is imperative to find ways to accelerate future iterations to keep pace with the new COVID-19 variants [3]. One of the ways to improve this process is to study how the population reacted to the first vaccination campaigns, the types of information/misinformation shared, and the impact this had on vaccination hesitancy.

Social media platforms are, of course, one of the main stages of this debate.

In the last years, microblogging services such as Twitter have seen an increase in popularity due to their immediacy and ease of use. Moreover, brands, institutional bodies, politicians, public figures, and traditional news outlets have realized the importance of having a presence on these platforms, which allow them to deliver messages with high impact and unprecedented reach [4,5].

The rapid spread of the pandemic, fast development of the vaccines, and increasing worries about their safety have been hot topics on social media since the very beginning.

The vaccination campaigns planned by national governments could therefore be seriously hampered by misinformation on such outlets [6,7]. Many recent studies [8] have taken great interest in analyzing different social media platforms to track the sentiment of users about COVID-19 vaccinations across different cities [9], looking for the main misconceptions and complaints about the COVID-19 control measures [10] and the confidence in the efficacy of the vaccines [11].

These are only few examples demonstrating why monitoring social media platforms is a highly informative and beneficial approach to discover health-related issues (eg, detecting mentions of adverse events [AEs]) and to better understand public opinion (eg, monitoring the information quality and contrasting the spread of fake news). From this point of view, modern systems for digital pharmacovigilance can deploy natural language processing techniques to collect and analyze online discussions. This allows for the identification of potential AEs that may not have been detected during clinical trials, enabling timely decisions to reduce their harm. In the near future, it is likely that even public health care systems will increase their monitoring activities on social media platforms, with the goal of identifying and treating health issues such as mental diseases, managing information by contrasting fake news, or launching prevention campaigns (eg, to mitigate vaccine hesitancy) [12].

Objective

We here present an overview of our system for monitoring and analyzing vaccine opinions. Its modules aim at generating insights from Twitter on the topic of COVID-19 vaccines. The tool collects tweets daily and analyzes them to extrapolate information about public reception of the vaccination campaigns on social media. The information on our interactive web portal is also broken down into easy-to-read charts for both specialized and general audiences. Figure 1 illustrates the architecture of the full system behind the web portal. The portal consists of a module dedicated to data collection and various modules dedicated to data processing. The main features of the system are: (1) Localization, (2) Hashtag Analysis, (3) News Sources Analysis, (4) Sentiment Analysis, and (5) Symptom Extraction.

The Symptom Extraction module, in particular, consists of a deep-learning architecture that we created specifically for this task, based on SpanBERT [13], an extension of the bidirectional encoder representations from transformers (BERT) model, which is one of the state-of-the-art models for AE detection [14-16].

Each processing module is built to extract specific information from the collected tweets (eg, the most used hashtag or the most shared links). This information is then cleaned and provided to the user through the web portal with interactive charts and diagrams. To ensure greater readability, colors and shapes were preferred over figures when presenting the data.

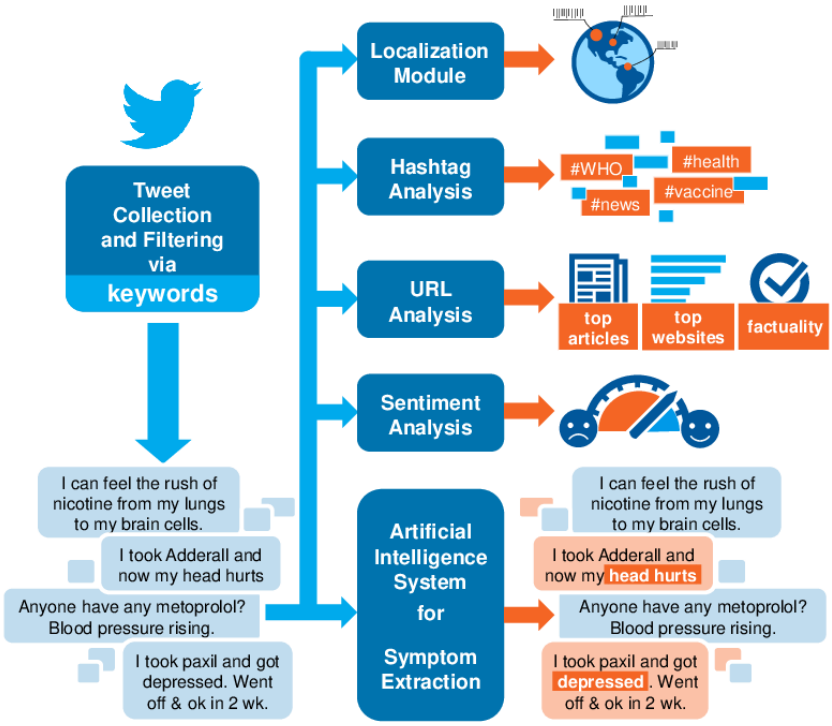
To summarize, our objective was to present a tool for the collection and processing of data on COVID-19 vaccines, followed by their visualization on a web dashboard [17].

In contrast to related previous works, we focused on monitoring tweets about specific vaccines. This allowed us to compare their public reception and how it changes over time. Besides combining various features that can be found separately in recent works, we also introduced innovative modules (eg, Symptom Extraction), which can offer new insights on the related public discourse.

The code for the data collection and the preprocessing tools, as well as all the precomputed statistics and the IDs of the tweets, can be openly accessed from GitHub [18]. The amount and type of data that can be shared openly are limited by Twitter's privacy policy. However, further information can be requested for research purposes. We also present a case study on the AstraZeneca vaccine, as an example of the analyses that can be carried out on the data using our system.

XSL•FO
RenderX

Figure 1. Schema of the full system architecture used to analyze the information displayed on the web portal.



Related Work

Since the start of the COVID-19 pandemic, organizations worldwide have stressed the need to collect and share all data available on the virus, its effects, and all related research [19]. As time passed, these resources grew in size, and some researchers also started analyzing data coming from social media.

For example, Kwok et al [10] collected 31,100 Australian tweets (from January 20, 2020, to October 22, 2020) related to COVID-19 vaccines. Their paper focuses on analyzing the sentiment and opinion of the users about the vaccines and the main recurring topics in the tweets. Similarly, Yan et al [9] collected and analyzed Reddit comments about COVID-19 vaccines from three Canadian cities (from July 13, 2020, to June 14, 2021), and performed a comparison of the sentiment and main discussion topics among the three locations. Other recent works focused on analyzing sentiment and discussion topics in tweets about COVID-19 generated in other countries and in different time periods [20-22].

These works were carried out on very specific time periods, which focused on a single aspect of the social media messages. A more comprehensive study was carried out on AvaxTweets [23], a public data set of Twitter posts and accounts that exhibited a strong stance against COVID-19 vaccines, collected between October 2020 and December 2020. The authors analyzed the accounts in terms of the most frequent hashtags, which news sources they shared, and their most likely political orientation, looking for useful insights on how to counter misinformation and vaccine hesitancy. However, both this and the preceding works were carried out on a limited time scale and aimed specifically at the research community, providing no tools or web interfaces to explore the data.

At the same time, various researchers focused not only on data collection but also on ways to start processing and visualizing the data to make them available for a broader public. COnVIDa [24] is a web-based platform that provides day-to-day interactive information on COVID-19–related conditions in Spain, collating data from various sources (eg, health databases, mortality reports, statistics, information on citizens' mobility from Google and Apple Maps). This project focuses on a single country and tries to combine different aspects of the situation to give the viewer a more complete visualization. CoVaxxy [25] is another data set and online dashboard that focuses on the correlations between tweets about COVID-19 vaccines, credibility of the shared news, and vaccine adoption on US geolocated posts. Sharma et al [26] presented another recent tool, which was used to collect and analyze Twitter conversations from March 1, 2020, to June 5, 2020. The dashboard visualizes sentiment information and trending topics, but focuses particularly on the credibility of the news shared in the tweets and on how misinformation spreads.

Our proposed system includes many of the features offered by these previous works, such as continuous day-to-day data collection and processing (since December 15, 2020), global data collection (not country-specific), sentiment analysis, and news sources analysis. Our tool differs from these previous works in relation to the following aspects: (1) focused monitoring of specific vaccines since the date of their approval, which enables users to compare the public's reaction to them; (2) a wide variety of processing modules (not focused on a single aspect) to provide a multifaced view of the social media discourse; (3) a comprehensive dashboard to visualize all of the processed data in an easy-to-read manner for different categories of users; (4) an innovative symptom extraction module to track the most discussed side effects; and (5) openly available code and data.

XSL•FO
RenderX

Methods

Data Collection

Tweets are collected using the Twitter application programming interface (API) [27]. To recover the most recent tweets mentioning a specific vaccine, we use the query "covid vaccine <vaccine_name>," where <vaccine_name> is the lowercase name of one of the monitored vaccines (originally Pfizer-BioNTech, AstraZeneca, and Moderna, which was then expanded to include the newly introduced vaccines). We require that all keywords are present in the tweet (either as text, hashtag, or as part of a link in the tweet) and that each query contains the name of only one vaccine.

Tweets are selected among the "most recent," as opposed to the "most popular," and retweets are discarded. This is done to avoid skewing the data with popular tweets produced by few influential users. Although we are collecting tweets in various languages, only those written in English are passed to the following stages of processing, as most of our current modules are language-dependent. Nonetheless, we are storing these data for future research, as we plan to overcome this limitation in the near future with the introduction of multilingual models (in particular for AE detection and sentiment analysis) and automated translation services. This will allow us to perform a complete analysis for all monitored languages.

The query is run every 24 hours, with a cap of 7000 requested tweets per day (to be divided among the monitored vaccines) imposed by the limits of the API. Despite the theoretical limitation, the number of new tweets that matched the query in the last 24 hours never exceeded 7000.

The body of the remaining messages undergoes additional preprocessing steps to identify possible duplicates and discard tweets that are practically identical (apart from hashtags, punctuation, or URLs). This situation occurs, for example, when users share a piece of news using the "Share on Twitter" button provided by news websites. If the user simply shares the news without adding any comments (or adding only a hashtag), the result is a high number of nearly identical tweets that do not provide additional information aside from the fact that the particular piece of news was shared multiple times. Such tweets are marked as "duplicated," but are not discarded because they can provide useful information on which articles went viral; nevertheless, they are marked to avoid introducing noise into other types of analyses.

Deduplication is performed by removing all hashtags, URLs, and punctuation, followed by (fuzzy) matching with the collection of "unique" tweets already collected.

Data collection started on December 10, 2020, concurrent with the Food and Drug Administration approval of the first COVID-19 vaccine (Pfizer-BioNTech), and the system has currently (September 7, 2021) analyzed over 650,000 tweets. Table 1 presents the names of the vaccines tracked at the time of writing and the date we started collecting related data.

Table 1. Names of the tracked vaccines and dates on which data collection started.

Vaccine name	Start date
Pfizer-BioNTech	December 10, 2020
AstraZeneca	December 11, 2020
Moderna	December 16, 2020
Sinopharm	February 24, 2021
Sputnik V	February 24, 2021
Sinovac	February 24, 2021
Johnson & Johnson	April 1, 2021

Ethics Considerations

Twitter is a major social network and, as such, has strict policies to regulate the ethical use of its data and the privacy of its users. Following their guidelines, we collect and store only the information needed for the processing steps that are currently implemented. We memorize the outputs of the modules and discard all of the sensitive data soon afterward. We also memorize the tweet ID, which allows us (and other researchers) to access the original tweet in the future, as long as the user does not delete it or change its visibility.

If a tweet needs to be displayed on a web interface, we use the API provided by Twitter, which allows us to display tweets on demand given their tweet ID (and only if their current visibility settings allow them to be displayed).

Data Processing of Incoming Data

Localization Module

The localization module enables tracking the geographical origin of the tweet, visualizing which countries are more involved in the discussion about the vaccines.

The geolocation is extracted directly from the tweet whenever possible. Users on Twitter can decide whether to share their location or not at any moment, and whether to geotag the places mentioned in their tweets. If the precise geolocation is not available, the module attempts to reconstruct it using the user's "location," a free-text field located in the user's profile. As such, "location" may contain imaginative terms or nonexistent locations (eg, "over the rainbow" or "the universe"). The module relies on heavy preprocessing, normalization, and cleaning steps to discard most of the noisy locations. The remaining locations

XSL•FO
RenderX

are passed on to Google Maps services [28] to determine the most accurate match.

The information is displayed on the web portal as a world map, where countries are shown in different shades of color; the larger the number of tweets coming from that country, the darker the color (the scale is exponential).

Hashtag Analysis

Hashtags are extracted from the most recent tweets only (the last 7 days, updated daily). We automatically remove a curated selection of hashtags, considered to be of low information content. In particular, we remove all hashtags containing the name of the vaccines that we are tracking (eg, #pfizer, #moderna, #biontech), words directly related to COVID-19 (eg, #covid, #coronavirus, #covidvaccine), and those containing the term "vaccine" only.

Information displayed on our web portal shows the hashtags as a colored treemap, where most of the tweeted hashtags cover a wider area and are darker in color.

News Sources Analysis

Sensitive topics such as health and vaccinations are fertile ground for the spread of misinformation, as proven by the amount of COVID-19–related fake news, which have been debunked in 2020 by fact-checking agencies (eg, PolitiFact [29]) and the precautions taken by the major social networks when dealing with posts mentioning the pandemic (eg, Facebook [30]).

An analysis of the most shared articles is of key importance to understand which sources of information are used by the public to inquire about vaccines.

We run the analysis by collecting all URLs contained in the tweets. We consider the most recent tweets only (last 7 days, updated daily) to reflect the impact of the most recent news. URLs are used both in their full form and considering their domain only. Unique URLs and domains are counted and used to provide two different kinds of information: the single most shared webpages (to individuate trending articles) and the most popular sources of information (intended as websites/domains, to individuate the favorite source of information in general).

Factuality Analysis

To further investigate the factuality of the URLs shared by users, we make use of Iffy+ [31], a website that provides an updated list of websites ranked by their factuality level. The lists provided by Iffy are the result of an aggregation of different popular fact-checking websites and trusted sources (eg, FactCheck.org, PolitiFact, and Wikipedia). The list we take into account is composed, for the most part, of websites with a low Media Bias/Fact Check (MBFC) factual level [32] and sources of fake news/misinformation identified by BuzzFeed, FactCheck.org, PolitiFact, and Wikipedia. We use this list to perform a factuality analysis over all of the collected tweets.

For each URL in a tweet, we check if its domain belongs to one of the websites on the Iffy+ list. If it does, we classify it according to its level of *MBFC factuality* (high, mixed, low, very low), and its *misinformation category* (eg, conspiracy, fake news). Factuality level and misinformation category might be not available for some of the websites ("not available"). If a domain is not part of the Iffy+ list, we assume it is a reliable ("reliable") source of information. All domains with a factuality level greater than or equal to "high" are labeled as "reliable." Only 0.0089% of the "reliable" URLs fall into this category.

We want to highlight that this analysis only explores the reliability of the links that the users are sharing, but not the legitimacy of the tweet as a whole. For example, a user might share a "fake news" article as a way to joke, mocking it in the text of the tweet. There might also be cases of users sharing links from reliable sources, accompanied by inflammatory or fake captions.

Sentiment Analysis

The sentiment analysis module aims at understanding the attitude of the users when sharing their opinions of the vaccines and their possible side effects. To understand the general sentiment of the crowd when talking about the vaccines, we employ a RoBERTa model [33] trained on tweets, which was fine-tuned for the sentiment analysis on the TweetEval Benchmark [34,35]. The model reached a macroaveraged recall of 72.6 (SD 0.4) on the test set.

This type of module is useful to interpret the general mood of the people speaking about the vaccines, about their possible side effects, or even about their vaccination experiences. In particular, this can be very effective to understand if a user is reporting facts, expressing distress, or expressing a positive attitude. For each tweet, the sentiment calculated using RoBERTa is normalized to a discrete set of values (positive, negative, or neutral) for ease of visualization.

Our web portal features an interactive line graph to observe how the sentiment varies in time. It allows the visitor to inspect the sentiment globally and compare the trends for the tweets mentioning specific vaccines.

Symptom Extraction

In the last decade, people have started discussing their personal health status on social media more and more often, looking for users with similar experiences, asking for suggestions, or reporting unexpected effects after the assumption of medicines. The latter represents an interesting type of information, as these effects might be considered as AE indicators for pharmacovigilance purposes.

Systems for the automatic extraction of AEs from informal and social media texts are at the core of a growing research trend in the field of natural language processing [36,37]. Moreover, several shared tasks have been recently organized within the audit command language community [38,39] to raise interest about this topic.

We evaluated different combinations of transformer-pretrained models and conditional random fields (CRFs) to create an effective deep-learning architecture for the task [16]. The best-performing model employs a neural network architecture based on SpanBERT [13] and CRFs [40], trained on the Adverse Event Detection data set of the Fourth Social Media Mining for Health Applications Shared Task (SMM4H) [41], thus

XSL•FO
RenderX

representing the current state of the art on the Shared Task [14,15] (Table 2).

These evaluation metrics resemble more closely how humans might perceive the correctness of the predictions. The AE extraction problem is modeled as token classification, tagging each word in the text as "inside" or "outside" of a symptom/AE.

The samples go through five main processing steps: text preprocessing, subword tokenization, BERT modeling, intermediate label prediction, CRF, final label aggregation.

Table 2. Performance of our adverse event extraction module against the previous top-performing models on the Fourth Social Media Mining for Health Applications Shared Task 2019.[a]

Architecture	Relaxed metrics[b]			Strict metrics		
	F1	Precision	Recall	F1	Precision	Recall
SpanBERT[c]+CRF[d] [15]	70.2	60.8	83.0	46.4	39.6	56.1
KFU [42]	65.8	55.4	81.0	46.4	38.9	57.9
THU_NGN [43]	65.3	61.4	69.7	35.6	32.8	38.8
MIDAS@IIITD [44]	64.1	53.7	79.3	32.8	27.4	40.9
TMRLeiden [45]	62.5	55.5	71.5	43.1	38.1	49.5

[a]Data were obtained from the public CodaLab leaderboard [46].

[b]Relaxed evaluation of the model's performances. A prediction that does not match exactly the correct adverse event, but overlaps with it (eg, "headache" instead of "strong headache") is not discarded but considered as a "partial match" (worth half a point).

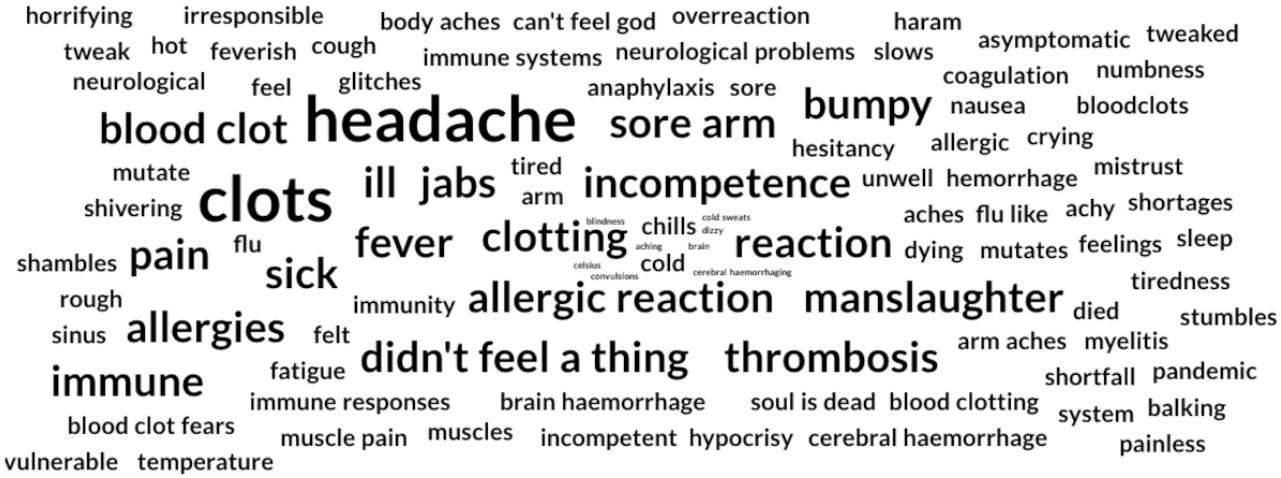
[c]BERT: bidirectional encoder representations from transformers.

[d]CRF: conditional random field.

The module of our system extracts all symptoms that are being discussed in the tweets. The data are then aggregated and visualized on the web portal as a word cloud. The data can be filtered by vaccine and by period of time to discover what concepts the crowd focused on at different stages of the vaccination campaign.

Figure 2 shows an example of the word cloud generated using tweets regarding the AstraZeneca vaccine following the thromboembolic events reported in several European countries during March 2021 [47].

Figure 2. Possible side effects of the AstraZeneca vaccine, as discussed on Twitter. The word cloud was generated using our adverse event extraction model and displayed on the web portal. The size of the words is proportional to their frequency.



Model Validation

The Sentiment Analysis and Symptom Extraction modules are based on deep-learning models, and it is thus crucial to verify their generalization capabilities outside benchmark environments. To more rigorously evaluate the performance of the modules mentioned above, we sampled and annotated a subset of the collected tweets to compare the model's predictions with human ground-truth labels on real-world data.

A total of 1000 tweets were extracted using stratified sampling to maintain the same distribution of tweets over months. Three annotators with high English proficiency (C1) were tasked to mark the sentiment of the tweets on a three-point scale (positive, neutral, negative) and highlight any vaccine-related AEs mentioned in them.

The gold sentiment of the tweet was decided by majority vote. The gold adverse events of the tweets were decided as the set of all sequences of words that were highlighted by at least 2 out

XSL•FO
RenderX

of 3 annotators. For example, if the annotations were "strong headache," "headache," and "having a strong headache," the final annotation would be "headache."

The human-generated annotations were used as ground truth to evaluate the performance of the two deep-learning modules on the real-world data and compare them with their performance on the benchmark data sets.

Results

Overall Results

First, we performed an initial analysis on the number of unique tweets and unique user accounts present in the collected data. As mentioned in the *Data Collection* subsection of the Methods, we took some precautions to avoid collecting duplicated data or skewing the data set by giving more weight to tweets posted by popular accounts. To verify if these strategies were successful, we inspected the ratio of unique tweets and users in the data set, month by month and overall.

Figure 3 shows the distribution of users depending on how many times their tweets appeared in the data set. We can clearly see a long-tail distribution, where 75% of the users only tweeted once, 92% of users tweeted at most three times, and 98% of users tweeted at most 10 times (ie, on average once per month). Looking at the users that tweeted more, most of them were news outlets, who tweeted from 50 to 578 times in the considered timespan (0.18% of the total users). The long-tail distribution is a good sign, as it shows that most of the users from whom we collected tweets are likely regular users and not influencers or content farms.

We then looked at the origin of the tweets that composed the data set. Figure 4 shows that 95% of the total tweets were posted by users that tweeted less than 100 times in the considered timeframe. This is another positive indication that the collection of tweets is not heavily influenced by a small number of super accounts, and thus the subsequent analysis should not suffer from this kind of bias.

Figure 3. Distribution of users depending on how many times they tweeted (the y axis is presented in logarithmic scale).

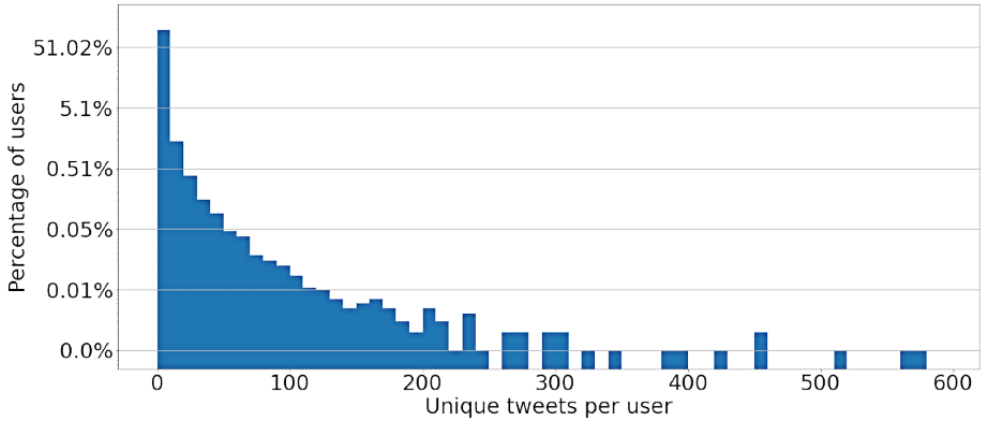
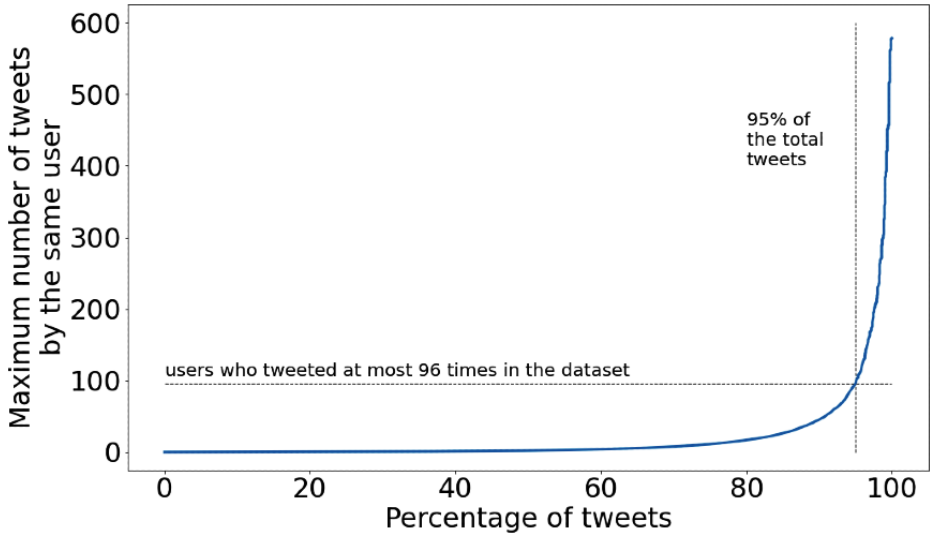


Figure 4. Percentage of tweets produced by a group of users, depending on how many tweets the user produced; 95% of the tweets in the data set are produced by users who tweeted at most 96 times in the considered timespan.



Finally, we calculated some statistics on a monthly basis, which are reported in Table 3. The mode and median were 1, confirming the findings discussed above. The average number of tweets per user remained stable at around 1.4 during the first months (December 2020 to March 2021). This number then increased to 1.5 in the period between April and June, following the start of the vaccination campaigns and the AstraZeneca

XSL•FO
RenderX

controversy (likely due to heightened news coverage). Following June, the average number of tweets per user went down again. The number of unique tweets and unique users considered each month was roughly stable.

Table 3. Statistics on the unique number of tweets and users for each month in the collected data set.

Month	Unique tweets, n	Unique users, n	Tweets per user			
			Maximum	Mean (SD)	Mode	Median
December 2020[a]	21,235	15,983	40	1.32 (1.29)	1	1
January 2021	42,891	30,294	71	1.42 (1.76)	1	1
February 2021	36,897	25,102	98	1.47 (1.98)	1	1
March 2021	51,469	35,402	181	1.45 (2.47)	1	1
April 2021	62,697	41,160	117	1.52 (2.45)	1	1
May 2021	48,785	32,263	134	1.51 (2.45)	1	1
June 2021	41,364	27,397	154	1.51 (2.45)	1	1
July 2021	42,742	29,371	139	1.46 (2.26)	1	1
August 2021	41,596	29,942	232	1.39 (2.09)	1	1
September 2021[a]	7064	5833	27	1.21 (0.84)	1	1
All	396,740	196011	578	2.02 (6.19)	1	1

[a]Partial data, not spanning the entirety of the month.

Localization

Since we are only considering English-language tweets, the most active countries were the United States, Canada, and the United Kingdom; followed by Nigeria, India, and Australia; and finally various European countries. Despite the language limitation that we imposed, the system detected tweets from almost all countries in the world.

We plan to remove the language limitation in the near future by means of the usage of automated translation services.

Hashtags

Most of the top hashtags were related to the concepts of "health," "news," or mentioned specific countries that made it to the top headlines due to recent outbreaks and similar accidents.

News Sources

The current data show a reassuring trend: the most popular sources of information are renowned newspapers (such as The New York Times or The Guardian), official institutional websites (eg, www.gov.uk), and scientific authorities (eg, the European Medicines Agency [EMA] and World Health Organization). It is also interesting to note that since the monitoring started in December 2020, the video-sharing platform YouTube has always been among the top-15 most shared domains. The top-5 most shared articles are displayed on the website as clickable links (displaying the URL and title of the page), while the 15 most popular domains are shown as a bar graph.

Factuality

The vast majority of the shared URLs were classified as having a "reliable" level of factuality (98%, see Figure 5). This seems to be confirmed if we look at the five most shared domains: theguardian.com (3.22%), nytimes.com (2.75%), reuters.com (2.40%), cnbc.com (1.77%), and abc.net.au (1.56%).

The remaining 2% was composed of domains classified mostly as *low* and *mixed* (ie, a website that is known to share both factual and nonfactual information). Figure 6 shows the factuality distribution of "unreliable" URLs (note that these data are presented on the logarithmic scale).

Looking at the misinformation categories for the "unreliable" domains (Figure 7), 49% were classified as "Conspiracy-Pseudoscience," 49% as generic "Fake-News" sources, and the remaining were subject to political biases.

XSL•FO
RenderX

Figure 5. Percentage of the Reliable and Unreliable URLs shared (y axis is presented in logarithmic scale).

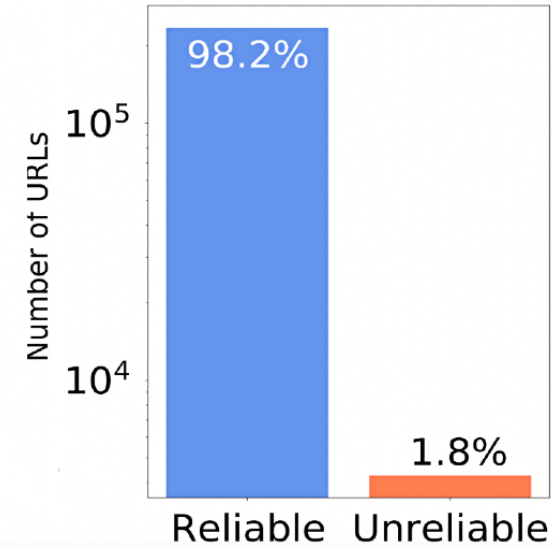


Figure 6. Distribution of Media Bias/Fact Check misinformation categories for "Unreliable" URLs. The y axis is presented in logarithmic scale. CP: Conspiracy-Pseudoscience; FN: Fake-News; N/A: Not Available; RC: Right-Center bias; R: Right bias; L: Left bias.

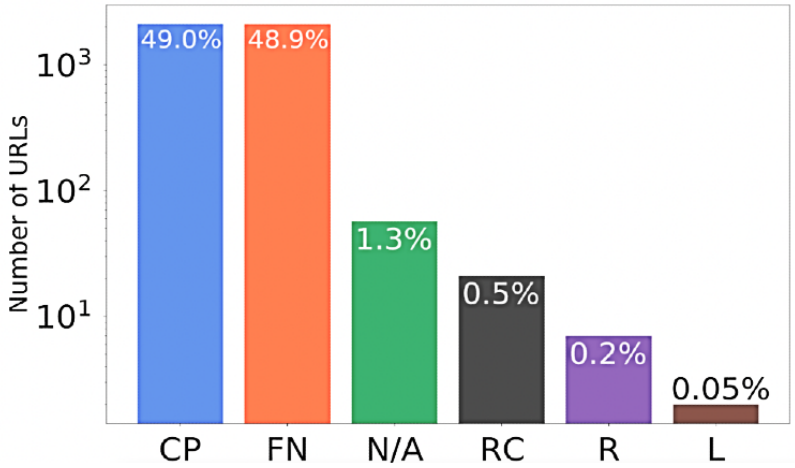
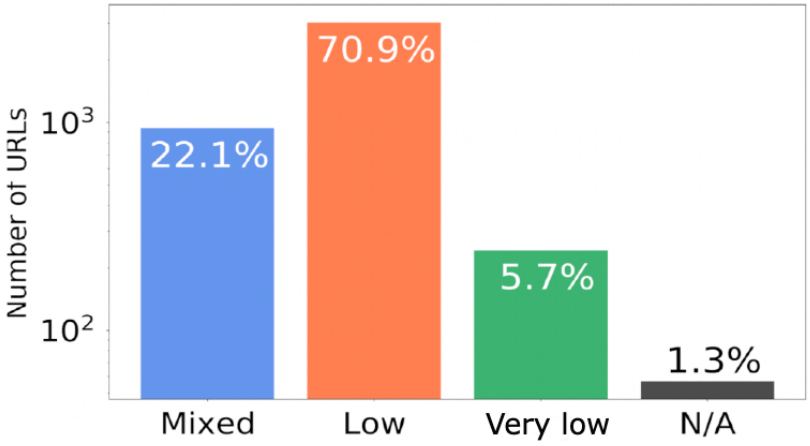


Figure 7. Distribution of Media Bias/Fact Check factuality level for "Unreliable" URLs. The y axis is presented in logarithmic scale. N/A: not applicable.



Sentiment Analysis

The global sentiment of the analyzed tweets was neutral/negative for most of the period of observation, with occasional spikes of positivity for individual vaccines. The negative trend might be enhanced by the fact that shocking, controversial, or tragic news tend to be shared and spread more easily on the internet when compared with other kinds of news.

Symptom Extraction

In the days preceding March 11, the most prominent concepts in AstraZeneca's word cloud were "headache" and "fever"; however, as soon as thromboembolic events started being

XSL•FO
RenderX

discussed on the internet, the system detected the shift in topic, and words such as “clots” and “thrombosis” quickly became noticeable in the cloud.

With regard to the other two vaccines, “allergic reactions,” “headache,” and “fever” were consistently among the most shared and discussed AEs. “Anaphylaxis” was one of the major concepts on Pfizer-BioNTech’s cloud for a long period of time at the beginning of the vaccination campaign, but is now slowly losing traction (this is evident in the word cloud on our web portal [17]).

This model could identify tweets containing potential AEs and highlight the mention of the symptoms. However, there are no mechanisms in place to verify the reliability of the tweets and there is no human fact-checking involved in the process. This means that, for the time being, there is virtually no distinction between symptoms that were actually reported by the users and exaggerations or hoaxes. This limitation is clearly stated on the web portal and the viewers are encouraged to further inspect the tweets on their own to have a clearer idea of what kind of messages lead to the prediction of the extracted symptoms. Clicking on any word in the word cloud displays a selection of the analyzed tweets that mentioned that concept in the selected time period.

The section “Evolution of mentioned symptoms over time” contains an analysis of the information that can be extracted by the representations produced by this module.

Finally, we would like to recall that the system was trained solely on the data provided during the SMM4H 2019 Shared Task. Even though it is one of the best performing models on this task, the model still suffers from the limitations of current AE extraction systems, such as the difficulty in making reliable distinctions between side effects (caused by medications), symptoms (caused by illnesses), and the names or descriptions of some medical conditions. For example, in the sentence “I have a *slipped vertebrae* and a *degenerative disk*,” the two medical conditions are identified as side effects by the system.

This is a common problem for such systems, which are often trained on data sets that are limited in size and linguistic variety.

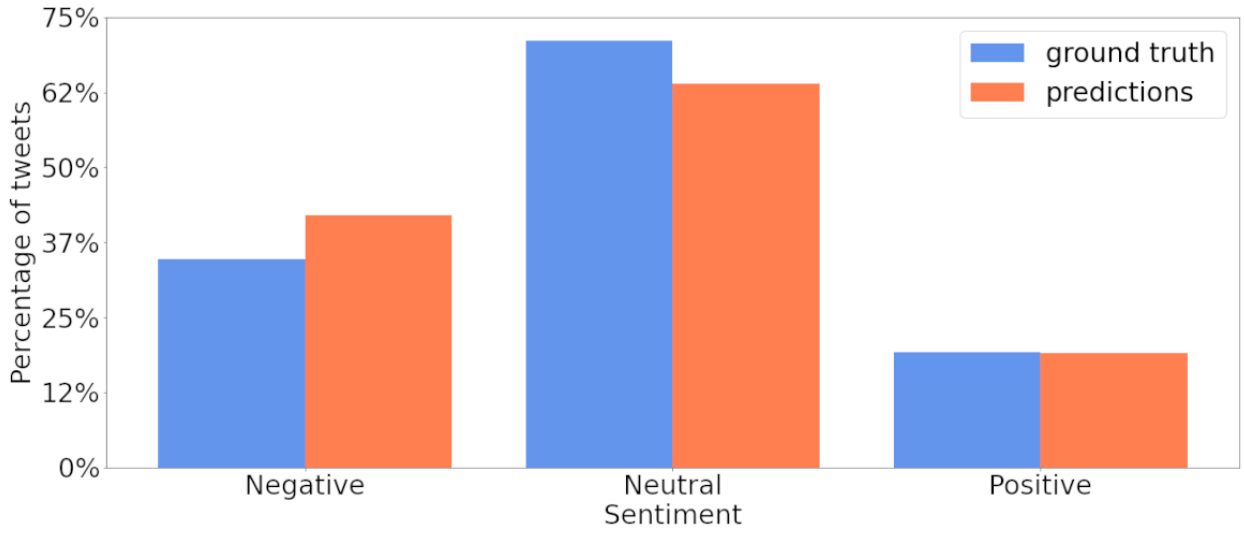
Model Validation

We experimentally evaluated the performance of both the Sentiment and Symptom Extraction modules using the subset of 1000 manually annotated tweets we created.

The performance of the Sentiment module on the real data was in line with that obtained on the benchmark data set, and its predictions were close to the ground truth. Figure 8 shows the sentiment distribution of the ground-truth labels (blue) and the predictions of the model (orange). The model leans slightly more toward negative sentiment. The performance (macroaveraged recall) on the subset of our data was 72.1. The model shows excellent generalization capabilities, which was in line with the performance recorded on the benchmark data set of 72.6 (SD 0.4).

To evaluate the Symptom Extraction module, we sampled our data set to have the same ratio of AE to no AE tweets as the benchmark data set SMM4H (57:43). The obtained relaxed F1 score was 63.3 (SD 0.7) (average over 10 sampling procedures), against 70.2 recorded on SMM4H. This gap in performance may be caused by the difference in the types of AEs present in the two data sets. For example, the benchmark data set focuses on sleep disorders and weight gain/loss, whereas the data we collected contain more instances of arm soreness and blood clotting, which the model had never encountered during training.

Figure 8. Comparison of the sentiment distributions of the manually annotated ground-truth labels (blue) and the model predictions (orange).



Case Study: AstraZeneca

Overview

To demonstrate the possible uses of our monitoring system as a research tool, we created a brief report regarding the AstraZeneca vaccine. In particular, we focused on analyzing the phenomenon of the alleged correlation between the vaccine and some specific side effects (eg, blood clots), in comparison with the other monitored vaccines.

XSL•FO
RenderX

Sentiment Trends for AstraZeneca

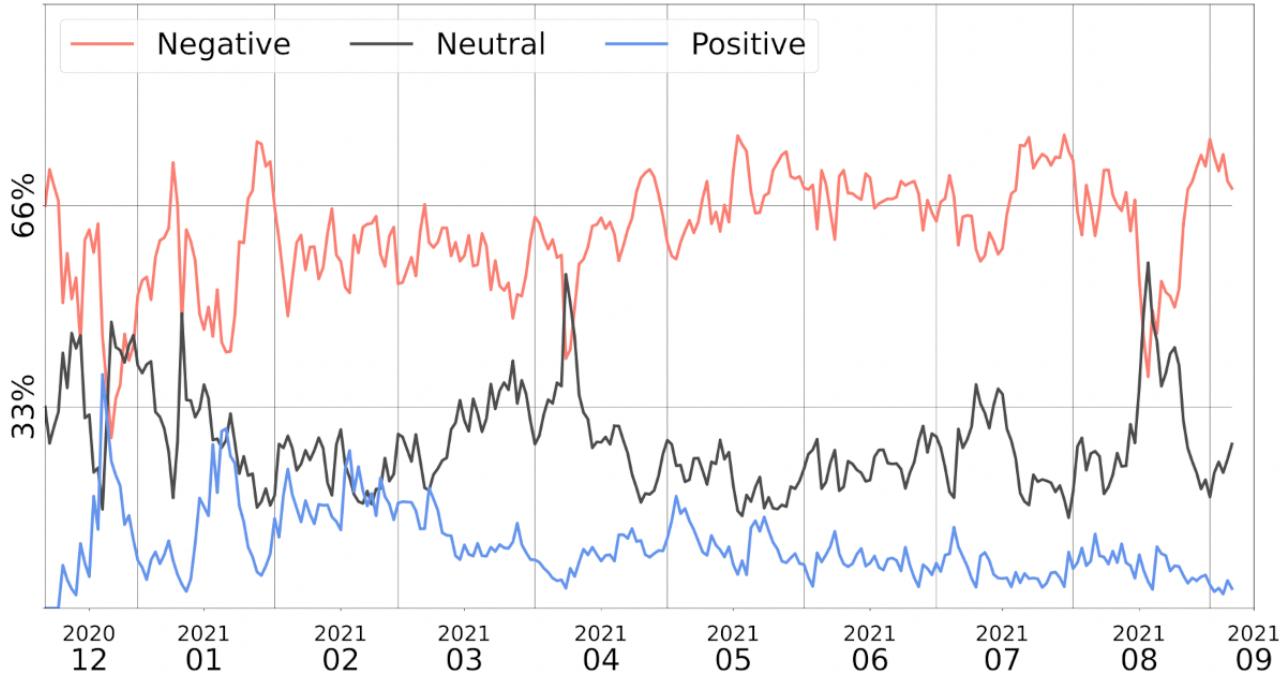
We start by providing a general overview of the sentiment of the crowd toward the vaccine, and how it varied in time. Figure 9 shows the day-by-day percentage of positive, neutral, and negative tweets about the AstraZeneca vaccine from the day the monitoring started (December 11, 2020) to the most recent date at the time of writing (early September 2021).

We can see that the sentiment toward the vaccine has been mostly negative for the entire time period. This is likely due to the tendency of negative and worrying topics or critical opinions to spread more easily on the internet. Approximately one third of the tweets were neutral, corresponding to people sharing factual information about the vaccine or showing neutrality and detachment toward the topic.

There was a noticeable trend of "nonnegativity" between December and January, when positive and neutral tweets covered more than half of the discussion.

This might be related to the publication of an important study [48] about the efficacy of the AstraZeneca vaccine and its approval by the EMA.

Figure 9. Monthly sentiment distribution in AstraZeneca vaccine–related tweets. The y-axis represents the percentage of negative (top, orange), neutral (middle, grey), and positive (bottom, blue) sentiment in the analyzed tweets. It is clear that the prevalent sentiment overall is "negative," but we can observe spikes of nonnegativity in December and January.



Mentions of Thromboembolic Events

We then compared the frequency with which Twitter users mentioned AEs related to "thrombosis" and "blood clotting" compared to other vaccine side effects.

Figure 10 shows the number of detected tweets for each day that contained clot-related AEs (red series) and any other AE (blue series).

The absolute number of tweets discussing AstraZeneca and its AEs increased from December 2020 to February 2021; however, blood clotting events were rarely discussed on Twitter.

This changed in the first half of March 2021, when the number of tweets discussing clot-related AEs had a peak. At that time, some European states (eg, Germany) stopped inoculations of the AstraZeneca vaccine due to the possible correlation between the clots and the vaccine, along with some suspicious deaths from ischemia.

Since then, the public attention on clot-related AEs has remained high and peaked periodically (see the red series), without losing track of the other topics (the number of tweets discussing other AEs remained high).

As specified above, not all tweets with clot-related references are AE reports: most of them come from people sharing or commenting news pieces about the vaccine.

We can also observe that in the last month, the chatter about AstraZeneca has diminished, as the blue and red series report less than 20 tweets per day.

Figure 11 offers a different perspective on the phenomenon: we collected all tweets mentioning blood clots and thrombosis, and divided them according to which vaccines they deal with. Before March 2021, most of the tweets dealing with clot-related AEs were associated with the Pfizer vaccine (75%-85%). With the wide news coverage about the cases related to AstraZeneca, the trend changed drastically, and over 80% of the tweets mentioning this kind of event were discussing AstraZeneca.

XSL•FO
RenderX

Figure 10. Number of tweets mentioning clot-related and nonclot-related keywords for the AstraZeneca vaccine (time is plotted on the x-axis and the number of tweets is plotted on the y-axis). The number of tweets mentioning clot-related adverse events (AEs) was initially next to zero, spiked in March 2021 due to media coverage, but has been gradually diminishing ever since. Tweets mentioning nonclot-related AEs show a more stable trend over time.

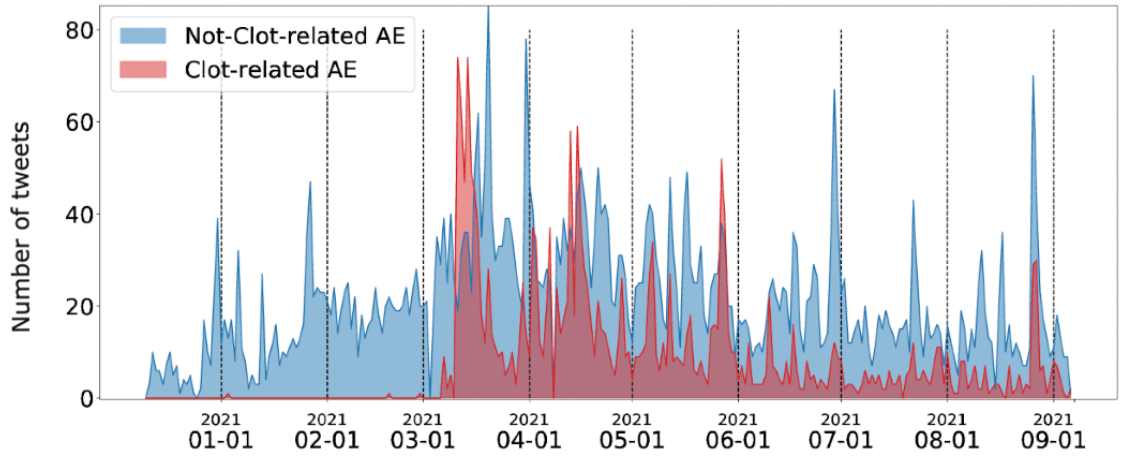
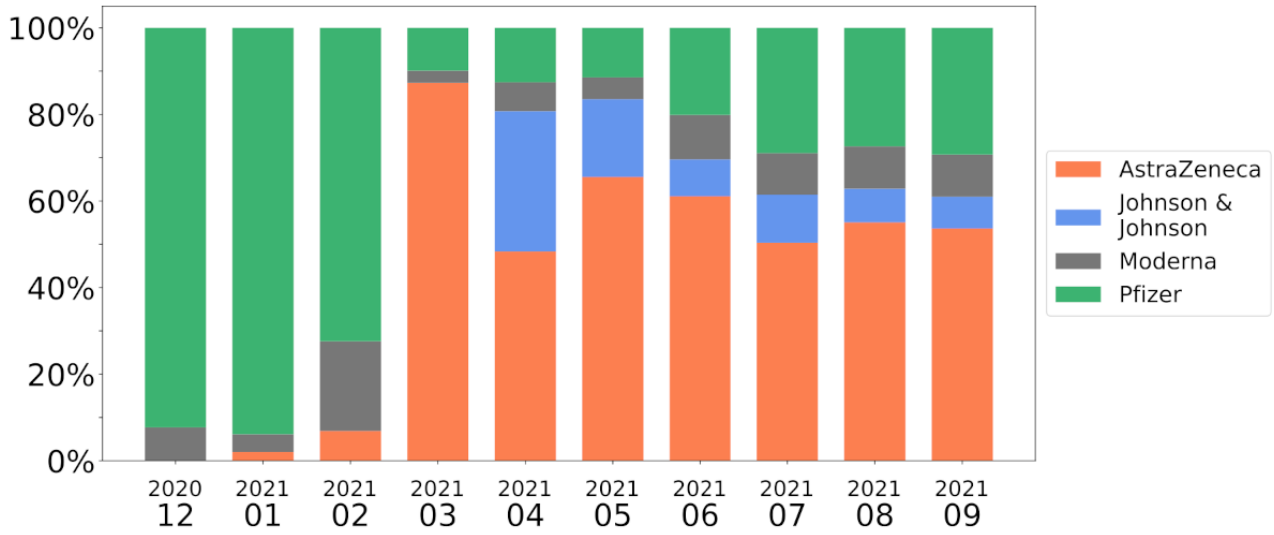


Figure 11. Monthly distribution of vaccine names mentioned in tweets with clot-related keywords (time on the x-axis, percentage of tweets on the y-axis). Most of the tweets were discussing clot-related adverse events connected to the Pfizer vaccine before March 2021, when the focus suddenly shifted to AstraZeneca.



Evolution of Mentioned Symptoms Over Time

The wide news coverage had a strong influence on the topics of discussion among Twitter users. This can be seen even more clearly in Figure 12, which shows three series of word clouds that represent how the main topics discussed on Twitter varied in time. The first row shows the most frequent AEs globally discussed (considering all tweets) for each month. The following rows show the evolution of the topics for the tweets that mention AstraZeneca, Moderna, or Pfizer only.

In the first 2 months (December 2020 to January 2021), all of the discussions were focused on widespread worries and doubts of the users (eg, allergies, neurological problems, immune responses).

During the following months, as the vaccination campaign proceeded, the focus slowly shifted toward the most common side effects that the vaccinated population was experiencing (eg, soreness at the arm, feeling sick, headache).

The news about AstraZeneca in March caused a dramatic shift of topic, not only in the tweets regarding that particular vaccine but also globally: the word “clot” suddenly appears in the global word cloud and becomes the most discussed topic for the following months (this also influences Pfizer’s word cloud, where the “clot” topic becomes slightly visible in April).

Looking at the latest available data, we can see that “blood clots” are still the most trending topic for AstraZeneca, but the global discussion has finally moved toward other topics such as “heart” problems. That said, if we look at all of the collected data, from December 2020 to September 2021, “clot” is the fourth most mentioned term globally (Figure 13), surpassed in popularity only by the broader concepts “arm,” “reaction,” and “sore.” This shows how great of an impact this episode had on social media.

XSL•FO
RenderX

Figure 12. Evolution of the global word cloud (top row, all vaccines included) and the specific word clouds of the following vaccines: AstraZeneca, Johnson & Johnson (J&J), Moderna, Pfizer. The suspected adverse events were extracted using our model.

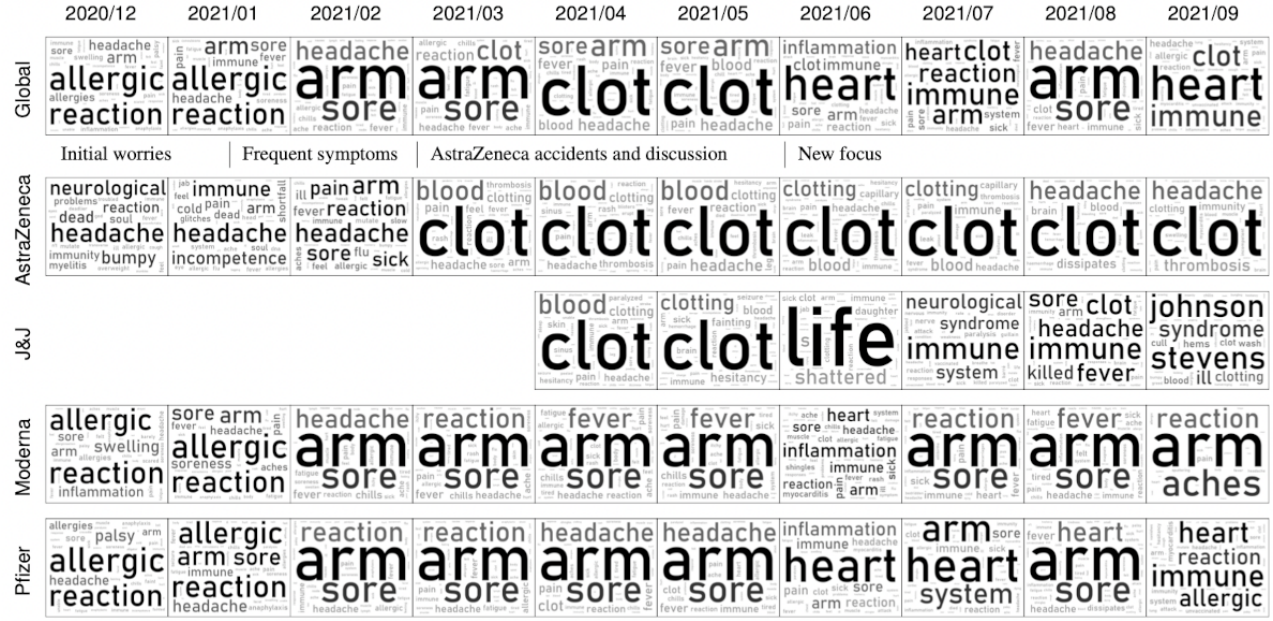


Figure 13. Top-5 most frequently mentioned terms globally and for the following vaccines: AstraZeneca, Johnson & Johnson (J&J), Moderna, Pfizer. This takes into consideration all of the collected tweets from December 15, 2020, to the beginning of September 2021.



Discussion

Intended Use Cases

Our web portal could be useful for different categories of users.

The first category is the general public. Owing to the intuitive interface and graphics, generic users can keep themselves up to date and be made aware of the kind of news that is circulating, what symptoms are being discussed for the various vaccines, and under which terms.

The second category is journalists and news outlets. The section of the web portal dedicated to news trends might provide insights for the press to better understand the digital audience and help in fighting misinformation. The other information might be interesting to explore to discover the latest most discussed topics.

The third category concerns users in the health care sector. The information on the most shared symptoms and possible AEs might be helpful to point the attention of the experts toward particular effects of the new vaccines.

Finally, scholars working in the field of biomedical natural language processing can benefit from the portal. The code of the AE extraction architecture is publicly available, and the web portal includes an explanatory page about the various implemented modules. The objective is to raise interest of the natural language processing community on this topic, and open the door to suggestions and possible collaborations.

Limitations

This project collects data from user-generated, unfiltered content, and makes use of automatic tools that have low and no human supervision. Therefore, it is important to highlight some limiting factors

The first limitation is the language barrier. As stated in the first sections, the current system is only able to analyze texts written in English. The COVID-19 vaccines are being distributed and discussed in several non-English–speaking countries, and therefore this data set is only a partial representation of public opinion. As stated in the *Data Collection* section, we plan to overcome this limitation with the use of multilingual models and/or automated translation services. We are already collecting tweets in other languages for the same time period, which will allow us to perform a complete comparative analysis in the future.

The second limitation relates to the demographics of Twitter users. Twitter is often used as a means to understand and monitor crowd opinions and real-world phenomena. However,

XSL•FO
RenderX

it is not always the case that Twitter users are a representative sample of the population of interest. A population can be examined along various axes (eg, age, geography, gender, ethnicity), and specific social media environments tend to overrepresent some sets of the population (eg, users coming from densely populated areas, higher level of education, higher income or computer literacy) [49,50].

Bias and misinformation spread on social media. Social media are also infamous for the creation of echo chambers [51], where users of the same mindset end up aggregating. This can "artificially" increase engagement with polarizing posts, which in turn become more visible and gain more weight in the analyses. Social media are highly polarizing environments, in which shocking, controversial, and generally "negative" posts are rewarded (and therefore can be found more frequently in the collected data) [52,53]. Our system tries to cope with this by handling data deduplication (removing viral copy-pasted tweets) and collecting the most recent tweets (as opposed to the most popular). This, however, does not remove the threats of echo chambers and misinformation. As future work, we plan to add a new module based on our previous work [54] to better analyze phenomena related to the spread of misinformation.

Finally, the correctness of deep-learning modules remains an inherent limitation. Both the Sentiment Analysis and Symptom Extraction modules are machine-learning modules, and as such can perform prediction errors with a known probability. If the data are shown to the public, users must be aware that they have to be taken with a grain of salt. This is why, on our dashboard, we make sure to include a disclaimer to warn the user about this issue whenever we display data produced by machine-learning algorithms.

Conclusions

We presented a tool connected with a web portal to monitor and display some key aspects of the public's reaction to COVID-19 vaccines.

The idea was born from the awareness that, in the current phase of the pandemic, it is of key importance to create tools to monitor reactions, opinions, doubts, and feedback of the population on the vaccines. Social media are a precious source of raw information, which can be exploited to gain insights for pharmacovigilance purposes (guiding the attention of health care experts on emerging effects) and help in fighting misinformation.

The system also provides an overview of the opinions of the Twittersphere through graphic representations to make them accessible to different categories of users.

One of the main features of this tool is the extraction of suspected AEs from tweets with a deep-learning model, which proved to be reactive to the shifts of topic in the internet chatter. A future improvement could be the extraction of AEs from tweets of different languages, using a multilingual model or an automated translation service.

All code, tweet IDs, and the precomputed statistics of the collected tweets are available at GitHub [18].

Conflicts of Interest

ES is a Senior Lead Data Scientist at Bayer Pharmaceuticals. The other authors have no conflicts of interest to declare.

References

1. Forni G, Mantovani A, COVID-19 Commission of Accademia Nazionale dei Lincei, Rome. COVID-19 vaccines: where we stand and challenges ahead. Cell Death Differ 2021 Feb;28(2):626-639 [FREE Full text] [doi: 10.1038/s41418-020-00720-9] [Medline: 33479399]
2. Goel RK, Nelson MA, Goel VY. COVID-19 vaccine rollout-scale and speed carry different implications for corruption. J Policy Model 2021;43(3):503-520 [FREE Full text] [doi: 10.1016/j.jpolmod.2021.04.003] [Medline: 33967361]
3. Sah P, Vilches TN, Moghadas SM, Fitzpatrick MC, Singer BH, Hotez PJ, et al. Accelerated vaccine rollout is imperative to mitigate highly transmissible COVID-19 variants. EClinicalMedicine 2021 May;35:100865 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100865] [Medline: 33937735]
4. Kruikemeier S. How political candidates use Twitter and the impact on votes. Comput Hum Behav 2014 May;34:131-139. [doi: 10.1016/j.chb.2014.01.025]
5. Jungherr A. Twitter use in election campaigns: a systematic literature review. J Inf Technol Politics 2015 Dec 21;13(1):72-91. [doi: 10.1080/19331681.2015.1132401]
6. Cinelli M, Quattrociocchi 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] [Medline: 33024152]
7. Sharevski F, Huff A, Jachim P, Pieroni E. (Mis)perceptions and engagement on Twitter: COVID-19 vaccine rumors on efficacy and mass immunization effort. Int J Inf Manag Data Insights 2022 Apr;2(1):100059. [doi: 10.1016/j.jjimei.2022.100059]
8. Karafillakis E, Martin S, Simas C, Olsson K, Takacs J, Dada S, et al. Methods for social media monitoring related to vaccination: systematic scoping review. JMIR Public Health Surveill 2021 Feb 08;7(2):e17149 [FREE Full text] [doi: 10.2196/17149] [Medline: 33555267]
9. Yan C, Law M, Nguyen S, Cheung J, Kong J. Comparing public sentiment toward COVID-19 vaccines across Canadian cities: analysis of comments on Reddit. J Med Internet Res 2021 Sep 24;23(9):e32685 [FREE Full text] [doi: 10.2196/32685] [Medline: 34519654]

XSL•FO
RenderX

10. 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] [Medline: 33886492]
11. 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] [Medline: 34623312]
12. Santus E, Marino N, Cirillo D, Chersoni E, Montagud A, Santuccione Chadha A, et al. Artificial intelligence-aided precision medicine for COVID-19: strategic areas of research and development. J Med Internet Res 2021 Mar 12;23(3):e22453 [FREE Full text] [doi: 10.2196/22453] [Medline: 33560998]
13. Joshi M, Chen D, Liu Y, Weld DS, Zettlemoyer L, Levy O. SpanBERT: improving pre-training by representing and predicting spans. Trans Assoc Comput Ling 2020 Dec;8:64-77. [doi: 10.1162/tacl_a_00300]
14. Portelli B, Passabì D, Lenzi E, Serra G, Santus E, Chersoni E. Improving adverse drug event extraction with SpanBERT on different text typologies. arXiv. URL: http://arxiv.org/abs/2105.08882 [accessed 2022-05-05]
15. Portelli B, Lenzi E, Chersoni E, Serra G, Santus E. BERT prescriptions to avoid unwanted headaches: a comparison of transformer architectures for adverse drug event detection. 2021 Presented at: 16th Conference of the European Chapter of the Association for Computational Linguistics; April 16, 2021; virtual. [doi: 10.18653/v1/2021.eacl-main.149]
16. Scaboro S, Portelli B, Chersoni E, Santus E, Serra G. NADE: a benchmark for robust adverse drug events extraction in face of negations. : Association for Computational Linguistics; 2021 Presented at: 2021 EMNLP Workshop W-NUT: The Seventh Workshop on Noisy User-generated Text; November 11, 2021; online p. 230-237. [doi: 10.18653/v1/2021.wnut-1.26]
17. COVID-19 Vaccine Opinion Analysis: Monitoring the Vaccines through Twitter Analysis. AI Lab, Università degli studi di Udine. URL: http://ailab.uniud.it/covid-vaccines/ [accessed 2022-05-05]
18. AilabUdineGit covid-vaccines-tools. GitHub. URL: https://github.com/AilabUdineGit/covid-vaccines-tools [accessed 2022-05-05]
19. Lynch CJ, Gore R. Short-range forecasting of COVID-19 during early onset at county, health district, and state geographic levels using seven methods: comparative forecasting study. J Med Internet Res 2021 Mar 23;23(3):e24925 [FREE Full text] [doi: 10.2196/24925] [Medline: 33621186]
20. Benis A, Chatsubi A, Levner E, Ashkenazi S. Change in threads on Twitter regarding influenza, vaccines, and vaccination during the COVID-19 pandemic: artificial intelligence-based infodemiology study. JMIR Infodemiology 2021 Oct 14;1(1):e31983 [FREE Full text] [doi: 10.2196/31983] [Medline: 34693212]
21. Zhang J, Wang Y, Shi M, Wang X. Factors driving the popularity and virality of COVID-19 vaccine discourse on Twitter: text mining and data visualization study. JMIR Public Health Surveill 2021 Dec 03;7(12):e32814 [FREE Full text] [doi: 10.2196/32814] [Medline: 34665761]
22. Liew TM, Lee CS. Examining the utility of social media in COVID-19 vaccination: unsupervised learning of 672,133 Twitter posts. JMIR Public Health Surveill 2021 Nov 03;7(11):e29789 [FREE Full text] [doi: 10.2196/29789] [Medline: 34583316]
23. Muric G, Wu Y, Ferrara E. COVID-19 vaccine hesitancy on social media: building a public Twitter data set of antivaccine content, vaccine misinformation, and conspiracies. JMIR Public Health Surveill 2021 Nov 17;7(11):e30642 [FREE Full text] [doi: 10.2196/30642] [Medline: 34653016]
24. Martínez Beltrán ET, Quiles Pérez M, Pastor-Galindo J, Nespoli P, García Clemente FJ, Gómez Mármol F. COnVIDa: COVID-19 multidisciplinary data collection and dashboard. J Biomed Inform 2021 May;117:103760 [FREE Full text] [doi: 10.1016/j.jbi.2021.103760] [Medline: 33798715]
25. DeVerna M, Pierri F, Truong B, Bollenbacher J, Axelrod D, Loynes N, et al. CoVaxxy: a collection of English-language Twitter posts about COVID-19 vaccines. arXiv. URL: http://arxiv.org/abs/2101.07694 [accessed 2022-05-05]
26. Sharma K, Seo S, Meng C, Rambhatla S, Liu Y. COVID-19 on Social Media: Analyzing Misinformation in Twitter Conversations. arXiv. 2020 Oct 22. URL: https://arxiv.org/abs/2003.12309 [accessed 2022-05-05]
27. Twitter API documentation. Twitter Inc. 2021. URL: https://developer.twitter.com/en/docs/twitter-api [accessed 2021-11-22]
28. Google Maps Platform. URL: https://developers.google.com/maps [accessed 2021-11-22]
29. Coronavirus. PolitiFact. 2020. URL: https://www.politifact.com/coronavirus/ [accessed 2021-04-18]
30. Jin KX. Keeping People Safe and Informed About the Coronavirus. Meta. 2020 Dec 18. URL: https://about.fb.com/news/2020/12/coronavirus/ [accessed 2021-01-13]
31. Iffy+ Mis/Disinfo Sites. Iffy. 2021. URL: https://iffy.news/iffy-plus/ [accessed 2021-09-20]
32. Media Bias Fact Check. URL: https://mediabiasfactcheck.com [accessed 2021-09-20]
33. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. 2019 Jul 26. URL: http://arxiv.org/abs/1907.11692 [accessed 2022-05-05]
34. Rosenthal S, Farra N, Nakov P. SemEval-2017 Task 4: Sentiment Analysis in Twitter. 2017 Presented at: 11th International Workshop on Semantic Evaluation (SemEval-2017); August, 2017; Vancouver, BC. [doi: 10.18653/v1/s17-2088]
35. Barbieri F, Camacho-Collados J, Espinosa AL, Neves L. TweetEval: unified benchmark and comparative evaluation for Tweet classification. 2020 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2020; November 2020; online p. 1644-1650. [doi: 10.18653/v1/2020.findings-emnlp.148]

XSL•FO
RenderX

36. Karimi S, Wang C, Metke-Jimenez A, Gaire R, Paris C. Text and data mining techniques in adverse drug reaction detection. ACM Comput Surv 2015 Jul 21;47(4):1-39. [doi: 10.1145/2719920]
37. Sarker A, Gonzalez G. Portable automatic text classification for adverse drug reaction detection via multi-corpus training. J Biomed Inform 2015 Feb;53:196-207 [FREE Full text] [doi: 10.1016/j.jbi.2014.11.002] [Medline: 25451103]
38. Weissenbacher D, Sarker A, Paul M, Gonzalez-Hernandez G. Overview of the Third Social Media Mining for Health (SMM4H) Shared Tasks at EMNLP. 2018 Presented at: 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task; October 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-5904]
39. Weissenbacher D, Sarker A, Magge A, Daughton A, O'Connor K, Paul M, et al. Overview of the Fourth Social Media Mining for Health (SMM4H) Shared Tasks at ACL 2019. 2019 Presented at: Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 21-30. [doi: 10.18653/v1/w19-3203]
40. Lafferty J, Mccallum A, Pereira F. Conditional random fields : probabilistic models for segmenting and labeling sequence data abstract. 2001 Presented at: ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning; June 28-July 1, 2001; Williamstown, MA p. 282-289. [doi: 10.1145/1015330.1015422]
41. Health Language Processing Lab. URL: https://healthlanguageprocessing.org/smm4h19/challenge [accessed 2020-03-22]
42. Miftahutdinov Z, Alimova I, Tutubalina E. KFU NLP Team at SMM4H 2019 Tasks: Want to Extract Adverse Drugs Reactions from Tweets? BERT to The Rescue. 2019 Presented at: Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 52-57. [doi: 10.18653/v1/w19-3207]
43. Ge S, Qi T, Wu C, Huang Y. Detecting and extracting of adverse drug reaction mentioning tweets with multi-head self attention. 2019 Presented at: Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 96-98. [doi: 10.18653/v1/w19-3214]
44. Mahata D, Anand S, Zhang H, Shahid S, Mehnaz L, Kumar Y, et al. MIDAS@SMM4H-2019: Identifying Adverse Drug Reactions and Personal Health Experience Mentions from Twitter. 2019 Presented at: Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 127-132. [doi: 10.18653/v1/w19-3223]
45. Dirkson A, Verberne S. Transfer learning for health-related Twitter data. 2019 Presented at: Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; August 2019; Florence, Italy p. 89-92. [doi: 10.18653/v1/w19-3212]
46. Sub-Task 2, ADR extraction, Post-Evaluation. CodaLab, CodaLab Competition – SMM4H'19 – Shared Task, 2020. URL: https://competitions.codalab.org/competitions/20798#results [accessed 2020-11-02]
47. COVID-19 Vaccine AstraZeneca: PRAC investigating cases of thromboembolic events - vaccine's benefits currently still outweigh risks - Update. European Medicines Agency. 2021 Mar 11. URL: https://www.ema.europa.eu/en/news/covid-19-vaccine-astrazeneca-prac-investigating-cases-thromboembolic-events-vaccines-benefits/ [accessed 2021-06-14]
48. Knoll MD, Wonodi C. Oxford-AstraZeneca COVID-19 vaccine efficacy. Lancet 2021 Jan 09;397(10269):72-74 [FREE Full text] [doi: 10.1016/S0140-6736(20)32623-4] [Medline: 33306990]
49. Smith A, Brenner J. Twitter use 2012. Pew Research Center. URL: https://www.pewresearch.org/internet/2012/05/31/twitter-use-2012/ [accessed 2022-01-14]
50. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J. Understanding the demographics of twitter users. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17-21, 2011; Barcelona, Spain p. 554 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14168
51. Usher N, Holcomb J, Littman J. Twitter makes it worse: political journalists, gendered echo chambers, and the amplification of gender bias. Int J Press Polit 2018 Jun 24;23(3):324-344. [doi: 10.1177/1940161218781254]
52. Fine JA, Hunt MF. Negativity and elite message diffusion on social media. Polit Behav 2021 Aug 11:1-20. [doi: 10.1007/s11109-021-09740-8]
53. Buder J, Rabl L, Feiks M, Badermann M, Zurstiege G. Does negatively toned language use on social media lead to attitude polarization? Comput Hum Behav 2021 Mar;116:106663. [doi: 10.31234/osf.io/dx9ws]
54. Portelli B, Zhao J, Schuster T, Serra G, Santus E. Distilling the evidence to augment fact verification models. 2020 Presented at: Third Workshop on Fact Extraction and VERification (FEVER); July 2020; online. [doi: 10.18653/v1/2020.fever-1.7]

Abbreviations

AE: adverse event
API: application programming interface
BERT: bidirectional encoder representations from transformers
CRF: conditional random field
EMA: European Medicines Agency
MBFC: Media Bias/Fact Check
SMM4H: Social Media Mining for Health Applications

XSL•FO
RenderX

Edited by C Basch; submitted 22.11.21; peer-reviewed by D Ceolin, R Gore; comments to author 13.12.21; revised version received 29.01.22; accepted 09.03.22; published 13.05.22.

Please cite as:
Portelli B, Scaboro S, Tonino R, Chersoni E, Santus E, Serra G
Monitoring User Opinions and Side Effects on COVID-19 Vaccines in the Twittersphere: Infodemiology Study of Tweets
J Med Internet Res 2022;24(5):e35115
URL: https://www.jmir.org/2022/5/e35115
doi:10.2196/35115
PMID:35446781

©Beatrice Portelli, Simone Scaboro, Roberto Tonino, Emmanuele Chersoni, Enrico Santus, Giuseppe Serra. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 13.05.2022. 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.

XSL•FO
RenderX

Original Paper

Multimodal Chronic Pain Therapy for Adults via Smartphone: Randomized Controlled Clinical Trial

Yolanda Morcillo-Muñoz[1*], BA, MSc, RN; Antonio José Sánchez-Guarnido[2*], BA, PhD; Silvia Calzón-Fernández[3*], BA, PhD; Isabel Baena-Parejo[4*], BA, PhD

[1]Primary Care, Andalusian Health Service, Algeciras, Spain
[2]Primary Care, Andalusian Health Service, Motril, Spain
[3]Primary Care, Andalusian Health Service, Sevilla, Spain
[4]Primary Care, Andalusian Health Service, Córdoba, Spain
[*]all authors contributed equally

Corresponding Author:
Yolanda Morcillo-Muñoz, BA, MSc, RN
Primary Care
Andalusian Health Service
Plaza de la Mujer s/n
District Campo de Gibraltar
Algeciras, 11207
Spain
Phone: 34 670946860
Email: ymmcadiz1969@gmail.com

Abstract

Background: Combination therapies delivered remotely via the internet or mobile devices are increasingly being used to improve and promote the self-management of chronic conditions. However, little is known regarding the long-term effects of these interventions.

Objective: The aim of this study is to evaluate the effectiveness of a multimodal intervention program that measures associated variables such as catastrophizing, pain acceptance, and quality of life using a mobile device in people with chronic pain in an outpatient setting.

Methods: A randomized controlled clinical trial was performed using parallel treatment groups. A total of 209 patients with chronic musculoskeletal pain were randomly assigned to one of the two study arms. The intervention group received a standard *web-based* psychosocial therapy-type program of activities through a smartphone for 6 weeks. The control group only had access to the *Find out more* section of the app, which contained audiovisual material for pain management based on a self-help approach. The primary outcome was catastrophizing measured using the Pain Catastrophizing Scale (PCS). Secondary outcomes were pain acceptance measured using the Chronic Pain Acceptance Questionnaire and health-related quality of life measured using the EuroQol Visual Analogue Scale. Assessments were conducted at baseline (T1), after treatment (T2), and at the 3-month follow-up (T3). The variations between the different phases were assessed using the percentage change rescaled with log base 2. The Cohen *d* was calculated based on the results of the linear mixed model. The investigators of the study who evaluated the results were not involved in patient recruitment and were blinded to the group assignment.

Results: Positive effects were found in the intervention group (T2–T1) in catastrophizing between the baseline and posttreatment phases (P<.001) and in helplessness (−0.72 vs 0.1; P=.002), rumination (−1.59 vs −0.53; P<.001), acceptance (0.38 vs 0.05; P=.001), and quality of life (0.43 vs −0.01; P=.002), although no significant changes were found for magnification (0.2 vs 0.77; P=.14) and satisfaction with health (0.25 vs −0.27; P=.13). Three months after treatment, significant differences were observed in the intervention group for the outcome variable of catastrophizing (PCS; −0.59 vs 0.2; P=.006) and the PCS subscales of helplessness (−0.65 vs 0.01; P=.07), rumination (1.23 vs −0.59; P=.04), and magnification (0.1 vs 0.86; P=.02).

Conclusions: The results of our study suggest that app-based mobile multidimensional treatments for adults with chronic pain improve catastrophizing, quality of life, and psychological flexibility immediately after treatment and that the effects are maintained for the primary outcome of catastrophizing for at least 3 months following treatment. Moreover, they promote self-management and can be used to complement face-to-face pain treatments.

XSL•FO
RenderX

Trial Registration: ClinicalTrials.gov NCT04509154; https://clinicaltrials.gov/ct2/show/NCT04509154

(J Med Internet Res 2022;24(5):e36114) doi:10.2196/36114

KEYWORDS

chronic pain; eHealth; multimodal intervention; catastrophizing; self-management; mHealth; mobile phone; randomized controlled trials

Introduction

Background

Pain is estimated to be among the top 10 conditions with the highest burden on health expenditure and health care resources and has a significant impact on patients' quality of life [1,2]. Pain is defined as an "unpleasant sensory and emotional experience associated with or similar to that associated with actual or potential tissue damage" [3]. When it affects one or more anatomical regions; persists for >3 months; and is associated with emotional (anxiety or depressed mood) and functional distress that interferes with work, social, and family life, pain is considered chronic [4]. Although not a frequent cause of mortality in itself, many people die experiencing pain, and even more people are living with pain [5,6]. Owing to the extremely high prevalence of chronic pain in the general population, it should be considered a health problem.

There is a large body of research on multidisciplinary treatment programs for adults with chronic pain, including reviews of the clinical evidence, effectiveness of pain treatments, and cost-effectiveness of chronic pain programs in outpatient settings [7,8]. A systematic review by Hauser et al [9] reported that multidisciplinary programs are effective in reducing chronic pain and improving patients' biopsychosocial situations and may also reduce the use of prescription medications. Similarly, the Canadian Agency for Drugs and Technologies in Health found that both multimodal therapy and monotherapy were beneficial for treating chronic pain. However, further research is needed to determine which type or combination of therapies can provide long-term benefits for these patients [8,10,11].

As chronic pain is a complex and multidimensional problem, it cannot be managed using medical therapies alone. Therefore, multidimensional treatments involving psychological therapies such as acceptance and commitment therapy (ACT), mindfulness, physical exercise, and health assets could play an important role in mitigating catastrophizing, improving pain acceptance, and reducing the use of psychotropic medications [8,12-14].

The available evidence supports the efficacy of several interventions for the self-management of chronic pain in outpatients. Some of the key components of these interventions are the administration of chronic pain medications according to the type of pain and patient comorbidities [7,10,14,15]; therapeutic exercise and patient education for the treatment of a wide range of musculoskeletal disorders [5,16,17]; patient education and counseling [18,19]; ACT as an evidence-based treatment for chronic pain intensity and depression [20,21]; mindfulness [22] to moderate the impact of catastrophizing on everyday pain [23,24]; and self-management [15,25,26]. Balancing activities with rest, stress management, emotion regulation, and appropriate physical exercise can also improve the quality of life of these patients [27]. An essential part of treatment is the early detection of catastrophic thinking, psychological inflexibility, and depression as the modification of these factors can reduce disability, decrease pain interference and intensity, and improve the ability of patients with chronic pain to perform activities of daily living [28].

In addition to face-to-face interventions for chronic conditions, interventions are increasingly being delivered via mobile apps or the internet as these technologies have become an essential part of people's daily lives and are always on call. In a systematic review of the use of information and communication technologies (ICTs) in chronically ill patients, 15.4% of the studies found that ICTs had a positive impact on patient empowerment or self-management, 14% showed an improvement in physical conditions and quality of life, and 5.1% reported greater self-efficacy for managing disease [29-32]. In the field of health, new technologies are being used for a variety of purposes, among them symptom assessment, psychoeducation, and treatment adherence. Such technologies have been shown to be beneficial for the provision of health care as they can improve patient accessibility and health care response, are instantaneous, and occur in real time [33-35].

Objectives

The aim of this study was to evaluate the effectiveness of a multimodal intervention program using a web-based smartphone or mobile device app. The app assesses pain perception by means of associated variables such as catastrophizing, pain acceptance, and quality of life in people with chronic pain.

The study was based on the hypothesis that participants assigned to the intervention group will exhibit less catastrophizing and emotional distress, more acceptance of pain, and improved ability to perform activities of daily living according to self-management values. Moreover, they will experience a better quality of life and fewer symptoms, pain intensity, anxiety, and depression. These outcomes will be measured after completion of the intervention period and at the 3-month follow-up and compared with a control group in line with the recommended outcomes in chronic pain research [36].

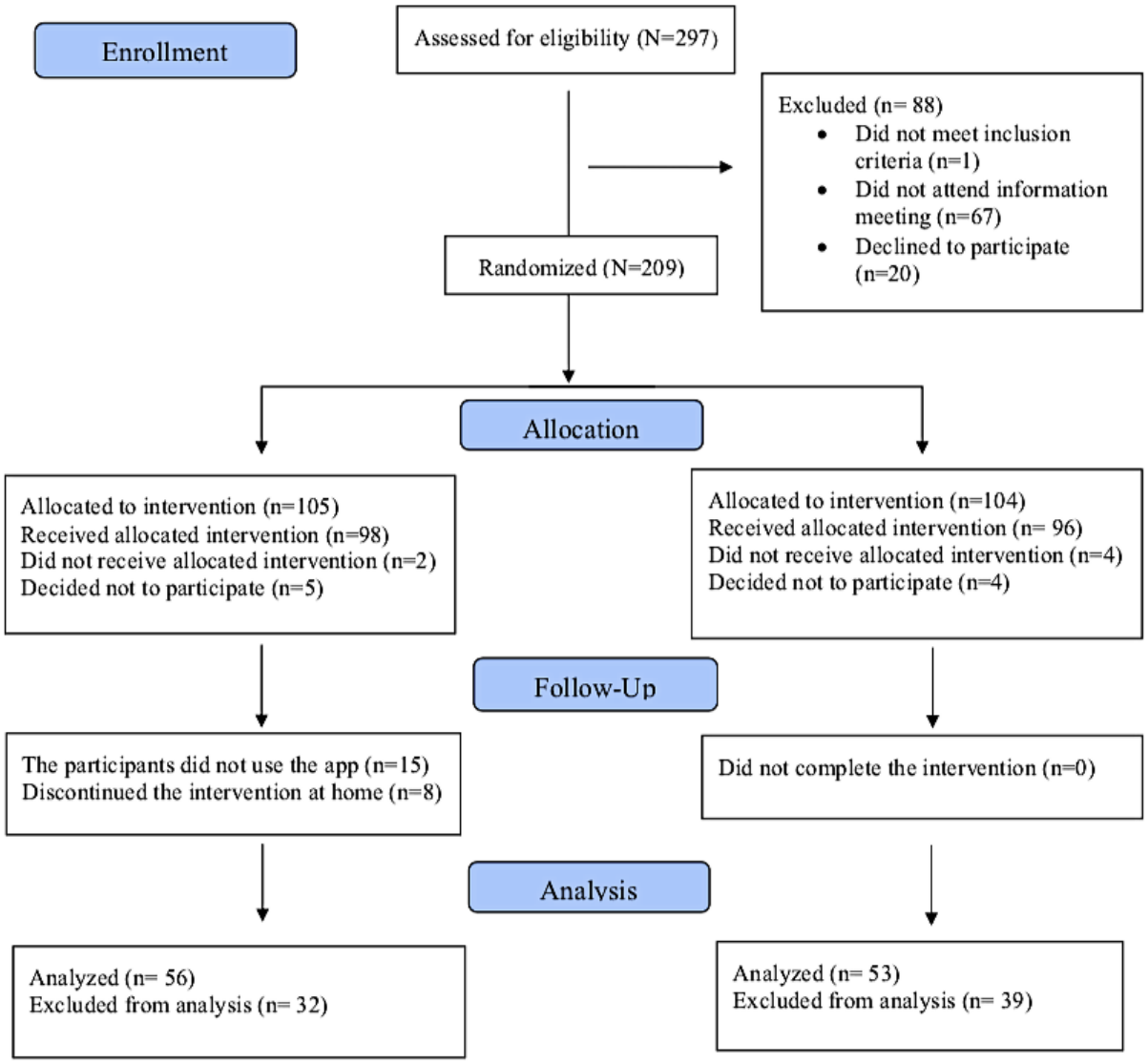
Methods

Study Design

A randomized controlled clinical trial was performed using parallel treatment groups [31,37]. Block randomization was used to ensure a similar number of participants in each group and in the intervention phase (Figure 1).

XSL•FO
RenderX

Figure 1. Flowchart of the recruitment process.



Mobile App–Based Multimodal Treatment

To develop the contents of the multimodal treatment for the mobile device app, a systematic literature review was carried out in a first phase following the Scottish Intercollegiate Guideline Network.

As part of a broader research program aimed at improving pain management, we first located articles that described the basic concepts of multi-professional treatments for chronic pain, provided recommendations, and included published evidence for the selected topics of interest. The purpose was to identify the characteristics of pharmacological and nonpharmacological interventions for people affected by chronic pain in different settings by means of a bibliographic search and reading of the literature, a synthesis of the results, and an assessment of the evidence. In the second part of our study, a participatory approach was used to prioritize self-management recommendations for chronic pain selected from the reviewed literature. Specifically, we designed a multimodal intervention protocol combining physical exercise, psychoeducational therapy, health assets, and pharmacological treatment that was delivered using a mobile device. The effectiveness of the intervention was subsequently evaluated by means of a clinical trial [38].

The interventions delivered in the mobile app include ACT and mindfulness exercises to promote greater pain acceptance, reduce the aversive component associated with pain, and help patients dispassionately recognize and observe both pain and related thoughts and emotions. Another group of activities aims to raise awareness of an individual's own values through a series of activities to recover a meaningful life project [22,23,39,40]. The exercise section provides tools and resources for patients to improve their physical, mental, and emotional well-being. The activities in this section include empowerment, stretching, relaxation, walking, and low-intensity exercises to help patients acquire good habits and learn about alternatives to improve their day-to-day life [5]. The activities in the pharmacological section aim to help patients better understand medications that reduce the intensity of pain. For each medication, the most common side effects and characteristics related to pain relief are described as well as which drugs are best suited to the patient's current health state, the risks of taking more than the recommended

XSL•FO
RenderX

dose of medication, and how to identify warning signs [41-43]. The activities in the health assets section are designed to improve patients' self-esteem and health by having them identify the individual, physical, institutional, associational, economic, and cultural assets and resources available in their community to help them cope better with situations of vulnerability and stress [44,45].

Development of the Mobile App

Our research team developed an app called NO+Dolor (NO+Pain) that contains an Android or iOS user interface. The app includes links to several multimedia resources (mainly audios and videos) and was designed based on game dynamics (gamification) to improve users' concentration, attention, and motivation. Indications on how to use the app and correctly perform the measures are provided in the instructions section of the app. Regarding the technical characteristics of the app, the information technology part required the design of a relational database implemented using MySQL Community Server 5.6. On the server side, a Java 1.8 communications application programming interface was implemented with representational state transfer architecture between the patients' mobile apps and the database using a client-server pattern. On the client side, hybrid mobile apps were implemented for Android and iOS operating systems. Apache Cordova 9.0 and jQuery Mobile 1.4.5 frameworks were integrated into the app, and HTML5 and CSS3 technologies were used for the presentation layer [46].

Participants

The study population comprised residents of municipalities belonging to the Cordoba South Health District of the Andalusian Health Service, Spain, who were registered in the health service user database. Patients were recruited from each of the district's 11 primary care centers by two collaborators: a nurse and a physician with experience in the follow-up of patients with chronic pain.

The sample was drawn from a database using Diraya electronic medical records. The inclusion criterion was being attended by primary care physicians and nurses. The database search was carried out on June 30, 2019, and identified 297 patients diagnosed with chronic musculoskeletal pain, of which 205 (69%) were women and 92 (31%) were men. Accepting an α error of .05 and a β error of .2 (statistical power of 80%) in a 2-tailed test and estimating 15% loss to follow-up, 296 participants were required to detect a statistically significant difference between 2 means at 3 points in our outcome variable with an estimated SD of 12.

All participants were under pharmacological treatment for chronic musculoskeletal pain (analgesics, anti-inflammatory drugs, antidepressants, anticonvulsants, or opioids) as previously indicated by their primary care physician and were asked to sign an informed consent form. The inclusion and exclusion criteria for the participants in the study are presented in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Patients aged ≥18 years with pain in any location
- Pain with a duration of ≥3 months
- Pain with an intensity of ≥4 on the Visual Numerical Scale
- Presenting one of the following characteristics: continuous pain or intermittent pain ≥5 days a week
- Able to use a smartphone
- Not participating in another research project

Exclusion criteria

- Cancer-related or postsurgical pain, patients with palliative care, and pediatric population
- Patients with acute pain (duration <3 months)
- History of brain injury
- Inability to complete the study forms because of mental disability or language barrier

Ethics Approval

This study was approved by the Cordoba Research Ethics Committee of the Andalusian Public Health System (CIF G-14825277; protocol code PI-0447-2017). Informed consent was approved by the Cordoba Research Ethics Committee and completed by the participants. This study is registered at ClinicalTrials.gov (NCT04509154).

Randomization and Blinding

Block randomization with a block size of 4 was used. The only stratification criterion was the reference health center of the patients. An automated recruitment form hosted on the REDCap (Research Electronic Data Capture; Vanderbilt University) platform of the Maimonides Biomedical Research Institute of Cordoba was used to randomize the patients by simply clicking a button. The data were transferred and recorded in an electronic notebook using the Data Entry Manager system. The statistician (Ipek Guler Caamaño and Juan Antonio Marín Sanz), principal investigator (YM-M), and coinvestigators (AJS-G, SC-F, and

XSL•FO
RenderX

MIB-P) of the study who evaluated the results were not involved in patient recruitment and were blinded to the group assignment. A total of 22 recruiters from the 11 primary care health centers of the Cordoba South Health District recruited the patients in 2019. They were also responsible for randomizing the patients (by clicking a button on the automated recruitment form) and were not blinded.

Treatment Procedures

All patients received a written invitation from their primary care physician or nurse to participate in the study. Two 8-hour face-to-face sessions were held at their reference health center led by a nurse and a primary care physician with experience in the follow-up of patients with chronic pain. At the group meeting, patients who agreed to participate voluntarily in the study were informed that they would receive instructions by email on how to download the mobile app with the treatment contents. The patients were also informed that, if they were selected to participate in the intervention group, the treatment would last from 6 to 8 weeks. The control group would only have access to the *Find out more* section of the app, which contained audiovisual materials for pain management from a self-help approach, such as information on the origin of chronic pain and advice for pain treatment and relaxation exercises.

The intervention group received the treatment via their smartphones for a period of 6 weeks after completing both face-to-face sessions. Pain intensity was measured daily on an 11-point numerical rating scale when the participants accessed the app.

To assess the treatments, self-reported questionnaires were sent to the participants by email and collected at three time points: upon admission to the program (T1), at week 6 of the intervention (T2), and at 3 months after the intervention (T3). The participants completed all 3 questionnaires at home and returned them by email.

Smartphone-Based Intervention

The intervention consisted of the implementation of a program of standard, interactive psychosocial therapy activities. The pain management app enables automatic monitoring, skill training, social support, education, goal setting, and achievement of four components: psychological wellness, exercise, pharmacological treatment, and health assets. Each week, the participants received 3 activities for each of the aforementioned components via the NO+Dolor app until completing the 6 weeks of treatment. All the activities were designed to be performed weekly except for the walking challenge, which was performed daily. The first time the participants completed an activity, they were awarded a star. The participants could perform the proposed activity as many times as they liked, but no more rewards were given until the following week. The more activities they completed, the more stars they were awarded, and the higher the percentage of goals reached by the patients each week (Multimedia Appendix 1). The app also had a *Consultation* section with a contact form where the participants could send any questions or comments. The form was then sent by email to the researchers so that they could respond to the inquiries.

Assessment Measures

The Spanish adaptation [47,48] of the Pain Catastrophizing Scale (PCS) [49] was used to measure the main outcome variable of the study [50]. The total score on the PCS is calculated by summing the responses to the 13 items and ranges from 0 to 52. The PCS subscales comprise three dimensions: (1) rumination, scored from 0 to 16 (difficulty inhibiting repetitive pain-related thoughts and inability to seek solutions); (2) magnification, scored from 0 to 12 (tendency to exaggerate distressing situations and negative aspects of pain and perception of oneself as unable to control pain); and (3) hopelessness or helplessness, scored from 0 to 24 (inability to cope effectively with pain). Higher scores indicate higher levels of catastrophizing. A score of ≥30 is considered a cutoff point for clinically significant catastrophizing levels. The Spanish version of the PCS has been shown to have adequate internal consistency (Cronbach α=.79), convergent validity and classificatory value, test-retest reliability (intraclass correlation coefficient=0.84), and sensitivity to change in size (effect size>2).

The Spanish adaptation [51] of the Chronic Pain Acceptance Questionnaire (CPAQ) [21] was used to measure engagement in life activities despite pain; willingness to experience pain without trying to control, change, or avoid it; ability to recognize the chronicity of pain; and the need to avoid or control pain. The CPAQ is a 20-item self-report questionnaire that rates pain acceptance on a 7-point Likert-type scale ranging from 0 (*never true*) to 6 (*always true*). The maximum possible score on the CPAQ is 120, with higher scores indicating higher pain acceptance. Initial studies on the acceptance and adaptation of the CPAQ have shown adequate internal consistency and expected correlations with measures of physical functioning and psychological distress. Subsequent studies have evaluated the content and dimensions of the questionnaire and identified two factors: activity engagement (Cronbach α=.82) and pain willingness (Cronbach α=.78) [21].

The Spanish adaptation of the EQ-5D [52] was used to measure health-related quality of life. This version can be used both in relatively healthy individuals (the general population) and in groups of patients with different conditions. Individuals assess their own health state first by level of severity in different dimensions (descriptive system) and then on the more general EuroQol Visual Analogue Scale (EQ-VAS) of 0 to 100 (*worst imaginable health state* and *best imaginable health state*, respectively). A third component of the EQ-5D is the social values index obtained for each health state generated by the instrument, which describes respondents' health state according to five dimensions: mobility, self-management, usual activities, pain or discomfort, and anxiety or depression. Regarding the instrument's psychometric properties, the test-retest reliability ranges from 0.86 to 0.90, and numerous studies have demonstrated its validity and sensitivity to change [53]. We included a question on subjective global improvement rated by the EQ-VAS from 0 to 100: *We would like you to indicate on this scale how good or bad your health state is today*. Pain intensity was measured using an 11-point numerical rating scale ranging from 0 (*no pain*) to 10 (*pain as bad as you can imagine*). The format of this rating was established in the latest Initiative on Methods, Measurement, and Pain Assessment in Clinical

XSL•FO
RenderX

Trials recommendations on core outcome measures for chronic pain clinical trials [36].

Data Analysis

A descriptive analysis was performed for the quantitative variables with mean and SD and for the qualitative variables with recounts (n) and proportions (%). Goodness-of-fit to a normal distribution was determined using the Shapiro-Wilk test, and homogeneity of variance was assessed using the Levene test. The quantitative variables of the treatment and control groups were compared using the Mann-Whitney *U* test, and the Pearson chi-square test was used to compare the qualitative variables and the Fisher-Freeman-Halton descriptive analysis. In addition, the variations between the different phases were assessed using the percentage change rescaled with log base 2. The association between the quantitative variables was determined using bivariate (Pearson linear correlation coefficient or Spearman ρ) and partial correlations controlling for the variables age and sex.

A linear mixed effects model [54,55] was subsequently used to assess changes over time for the repeated measurements of the pain questionnaire scores at 3 time points between the control and treatment groups. Linear mixed effects models account for variability between participants and between repeated measurements in the same participant simultaneously. To obtain different trajectories for each group (experimental vs control) over time, we included the intercept and slope effect as random effects and time, group, and the interaction term (group×time) as fixed effects. The variance-covariance structure was fixed to an unstructured matrix, and the random effects and error terms were assumed to have a normal distribution. Furthermore, the Cohen *d* was calculated based on the results of the linear mixed model. The R project *nlme* package (version 3.5.0; R Foundation for Statistical Computing) was used to estimate all the regression models. The established level of statistical significance was $P<.05$.

Results

Participants

A total of 297 participants (n=205, 69% women and n=92, 31% men) were initially invited to participate in the study. Of these 297 participants, 1 (0.3%) was excluded for not meeting the inclusion criteria, 67 (22.6%) were excluded because they did not attend the meeting, and 20 (6.7%) were excluded because they declined to participate. A total of 209 participants were randomized into 1 of the 2 study arms. After randomization, 2.4% (5/209) of the participants were excluded because they declined to participate, dropped out before the intervention, or did not know how to use the technologies. In the intervention group, 15% (15/98) of the patients did not complete the intervention. The analysis was performed on patients who completed all 3 questionnaires at baseline, upon completion of the intervention, and 3 months after the intervention (Figure 1).

Table 1 shows the demographic data and baseline characteristics of the sample by group. The outcome variables showed measures for PCS ($P=.20$), CPAQ ($P=.07$), and EQ-5D ($P=.26$) at the beginning of the intervention (baseline), in which participants in both groups did not differ in pain catastrophizing, quality of life, or pain acceptance. No differences were found in sociodemographic variables or use of medications (Table 1).

XSL•FO
RenderX

Table 1. Characteristics of the sample (N=194).

Characteristic	Intervention group (n=98)	Control group (n=96)	*P* value
Age (years), mean (SD)	51.2 (11.2)	50.3 (10.2)	.99[a]
Gender, n (%)			.99
Female	78 (80)	77 (80)	
Male	20 (20)	19 (20)	
Marital status, n (%)			.59
Married	81 (83)	78 (81)	
Divorced	6 (6)	8 (8)	
Single	6 (6)	6 (6)	
Widowed	2 (2)	4 (4)	
Employment status, n (%)			.27
Unemployed	17 (17)	15 (16)	
Employed full-time	27 (28)	29 (30)	
Employed part-time	13 (13)	3 (3)	
Disability	11 (11)	12 (13)	
Home care	22 (22)	28 (29)	
Other combination of the aforementioned characteristics or unknown	8 (8)	9 (9)	
Level of education, n (%)			.16
No schooling	13 (13)	7 (7)	
8 to 10 years (elementary)	41 (42)	54 (56)	
10 to 12 years (high school)	36 (37)	26 (27)	
>12 years (higher education)	8 (8)	9 (9)	
Taking medications, n (%)			.68
Yes	96 (98)	92 (96)	
No	2 (2)	4 (4)	
Annual income, € (US $), n (%)			.45
<10,000 (10,857)	11 (11)	12 (26)	
10,000 to 20,000 (10,857 to 21,714)	73 (74)	74 (77)	
>20,000 (21,714)	14 (14)	10 (5)	
EQ-5D and pain, n (%)	56 (100)	53 (100)	.34
Moderate pain	34 (61)	27 (51)	
Extreme pain	22 (39)	26 (49)	
EQ-5D and anxiety and depression, n (%)	56 (100)	52 (100)	.90
I am not anxious or depressed	13 (23)	13 (25)	
I am moderately anxious or depressed	32 (57)	30 (58)	
I am extremely anxious or depressed	11 (20)	9 (17)	
PCS[b], mean (SD)	29.86 (13.27)[c]	27.7 (12.77)[d]	.20
CPAQ[e], mean (SD)	64.16 (18.89)[d]	66.77 (18.4)[f]	.07
EQ-5D[g], mean (SD)	0.45 (0.19)[h]	0.43 (0.21)[i]	.26
EQ-VAS[j], mean (SD)	48.22 (18.74)[k]	44.23 (23.49)[l]	.16

[a]n=194.

[b]PCS: Pain Catastrophizing Scale (score 0-52).

XSL•FO
RenderX

[c]n=55.
[d]n=54.
[e]CPAQ: Chronic Pain Acceptance Questionnaire (score 0-120).
[f]n=52.
[g]EQ-5D score −0.654 to 1.
[h]n=53.
[i]n=51.
[j]EQ-VAS: EuroQol Visual Analogue Scale (0-100).
[k]n=50.
[l]n=47.

Primary Outcome: Catastrophizing

Overview

The descriptive results for the differences between the control and intervention groups regarding measures of catastrophizing at baseline, upon completion of the treatment, and during follow-up are presented in Table 2. The between-group variations over time for the primary outcome variables are presented in Table 3.

XSL•FO
RenderX

Table 2. Median, mean, SD, and differences between groups for the primary outcome measure at admission to the program (T1), immediately after treatment (T2), and at the 3-month follow-up (T3) for completers.

Primary outcome measure (scale)	T1[a]			T2[a]			T3[a]		
	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value
PCS[b]			.20			.001			.09
Control	26.5 (17.25-38.5)	27.7 (12.77)[c]		32.5 (23.75-42.0)	31.82 (12.06)[d]		31.0 (22.5-43.0)	31.41 (12.61)[e]	
Intervention	29.0 (22.0-39.0)	29.86 (13.27)[f]		19.5 (14.25-28.25)	20.86 (11.25)[g]		27.0 (17.0-35.0)	25.78 (14.12)[h]	
Helplessness			.31			.001			.07
Control	12.5 (8.0-16.25)	12.48 (5.8)[i]		14.0 (10.5-19.0)	14.22 (5.82)[e]		14.0 (9.75-18.75)	13.65 (5.75)[j]	
Intervention	14.0 (8.25-18.0)	13.09 (6.45)[c]		8.5 (5.25-12.0)	8.91 (5.38)[g]		10.0 (6.0-17.0)	11.09 (6.48)[g]	
Rumination			.14			.004			.12
Control	10.0 (6.0-12.0)	9.19 (4.16)[k]		7.0 (4.0-8.25)	6.32 (2.99)[d]		6.0 (4.0-9.0)	6.28 (2.84)[l]	
Intervention	10.0 (7.5-14.0)	10.11 (4.11)[m]		4.0 (2.0-5.0)	4.05 (2.4)[n]		5.0 (3.0-7.0)	5.24 (3.16)[n]	
Magnification			.20			.007			.12
Control	6.5 (4.0-8.75)	6.57 (3.17)[c]		13.0 (8.75-14.25)	11.29 (4.28)[d]		12.0 (7.5-15.5)	11.33 (4.56)[e]	
Intervention	7.0 (4.25-9.75)	7.06 (3.22)[c]		8.0 (6.0-12.0)	8.29 (4.14)[n]		10.5 (5.75-12.0)	9.55 (5.04)[g]	

[a]At admission to the program (T1), at the end of the intervention at 6 weeks (T2), and at the 3-month follow-up (T3).

[b]PCS: Pain Catastrophizing Scale.

[c]n=54.

[d]n=28.

[e]n=27.

[f]n=56.

[g]n=22.

[h]n=23.

[i]n=52.

[j]n=26.

[k]n=53.

[l]n=25.

[m]n=55.

[n]n=21.

XSL•FO
RenderX

Table 3. Between-group differences and changes over time for the primary outcome measure and Pain Catastrophizing Scale (PCS) subscales.

Primary outcome measure (scale)	T2–T1[a]			T3–T2[a]			T3–T1[a]		
	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value
PCS			<.001			.24			.006
Control	0.05 (−0.11 to 0.73)	0.29 (1.0)[b]		−0.03 (−0.19 to 0.21)	−0.07 (0.32)[c]		0.1 (−0.14 to 0.36)	0.21 (1.13)[d]	
Intervention	−0.5 (−1.0 to −0.14)	−0.67 (0.72)[e]		0.0 (−0.27 to 0.58)	0.02 (0.65)[f]		−0.15 (−1.0 to 0.08)	−0.59 (0.91)[e]	
Helplessness			.002			.21			.007
Control	0.0 (−0.23 to 0.4)	0.1 (0.84)[g]		−0.07 (−0.41 to 0.04)	−0.11 (0.33)[h]		0.0 (−0.32 to 0.48)	0.01 (0.88)[i]	
Intervention	−0.79 (−1.14 to −0.07)	−0.72 (0.78)[j]		0.0 (−0.35 to 0.44)	0.11 (0.8)[f]		−0.33 (−1.06 to −0.08)	−0.65 (0.92)[j]	
Rumination			<.001			.48			.049
Control	−0.58 (−0.83 to 0.0)	−0.53 (0.89)[k]		0.0 (0.0 to 0.39)	0.29 (0.86)[h]		−0.71 (−1.0 to −0.26)	−0.59 (0.96)[i]	
Intervention	−1.42 (−2.27 to −0.79)	−1.59 (0.95)[l]		0.0 (0.0 to 0.57)	0.22 (0.55)[h]		−0.85 (−1.82 to −0.58)	−1.23 (0.84)[l]	
Magnification			.14			.41			.03
Control	0.56 (0.32 to 0.97)	0.77 (1.0)[b]		0.12 (−0.14 to 0.22)	−0.02 (0.48)[c]		0.68 (0.49 to 1.25)	0.86 (1.02)[d]	
Intervention	0.42 (0.29 to 0.66)	0.2 (1.02)[l]		0.0 (−0.13 to 0.46)	−0.03 (0.83)[f]		0.58 (−0.43 to 0.68)	0.1 (1.16)[j]	

[a]At admission to the program (T1), at the end of the intervention at 6 weeks (T2), and at the 3-month follow-up (T3).
[b]n=26.
[c]n=16.
[d]n=27.
[e]n=21.
[f]n=17.
[g]n=23.
[h]n=15.
[i]n=24.
[j]n=20.
[k]n=25.
[l]n=19.

Between-Group Effects

Immediately following the treatment (T2), statistically significant improvements were observed in the intervention group for catastrophizing (20.86 vs 31.82; *P*=.001) and the catastrophizing subscales of helplessness (8.91 vs 14.22; *P*=.001), rumination (4.05 vs 6.32; *P*=.004), and magnification (8.29 vs 11.29; *P*=.007). However, at 3 months of follow-up, the mean remained lower (25.78 vs 31.41; *P*=.09), although it was not statistically significant (Table 2).

Within-Group Effects

Positive effects were observed at the different treatment times according to the percentage change rescaled by log base 2. Specifically, positive effects were found for the intervention group (T2–T1) in catastrophizing between the baseline and posttreatment phases (*P*<.001) and in the subscales of helplessness (−0.72 vs 0.1; *P*=.002) and rumination (−1.59 vs −0.53; *P*<.001), although no significant changes were found for magnification (0.2 vs 0.77; *P*=.14). Significant results were also found for catastrophizing at the 3-month follow-up with respect to the baseline (−0.59 vs 0.2; *P*=.006) and the subscales of helplessness (−0.65 vs 0.01; *P*=.07), rumination (−1.23 vs −0.59; *P*=.04), and magnification (0.1 vs 0.86; *P*=.02), all of which improved 3 months after completing treatment (Table 3).

We also measured changes in the pain questionnaire scores between the control and treatment groups at the 3 time points by applying a linear mixed effects model. Statistical differences were found only between the 2 groups for changes in the PCS score over time. In addition, the interaction effect between time and the intervention group was −6.47 (*P*=.001), thus indicating a significant decrease in PCS scores over time in the intervention group compared with the control group (Table 4).

XSL•FO
RenderX

Table 4. Changes between groups in Pain Catastrophizing Scale scores over time.

Coefficient	Values, mean (SD)	*P* value
Intercept	25.747 (2.572)	<.001
Time	2.479 (1.336)	.07
Intervention (reference: control)	7.528 (3.669)	.04
Time[a] intervention (reference: time control)	−6.476 (1.991)	.001

[a]Pain Catastrophizing Scale scores over time in the intervention group compared with the control group.

Secondary Outcomes

Overview

Table 5 shows the results for the variables of acceptance (CPAQ), quality of life (EQ-5D), and overall health state (EQ-VAS), whereas Table 6 shows variations over time for the secondary outcomes.

Table 5. Mean and SD for the secondary outcomes at admission (T1), immediately after treatment (T2), and 3 months after the intervention period (T3) for completers.

Secondary outcome measure (scale)	T1[a]			T2[a]			T3[a]		
	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value
CPAQ[b]			.07			.14			.47
Control	68.0 (58.0-78.0)	66.77 (18.4)[c]		65.5 (52.75-72.5)	63.82 (12.47)[d]		65.0 (54.5-71.5)	65.67 (16.06)[e]	
Intervention	63.0 (53.75-71.0)	64.16 (18.89)[f]		67.0 (59.0-78.0)	68.23 (14.43)[g]		62.0 (54.5-73.0)	64.48 (21.76)[h]	
EQ-5D			.27			.008			.30
Control	0.46 (0.22-0.59)	0.43 (0.21)[c]		0.36 (0.22-0.56)	0.41 (0.22)[d]		0.37 (0.22-0.52)	0.39 (0.19)[e]	
Intervention	0.48 (0.23-0.59)	0.45 (0.19)[f]		0.56 (0.47-0.72)	0.55 (0.17)[g]		0.41 (0.22-0.59)	0.43 (0.2)[h]	
EQ-VAS[i]			.16			.02			.03
Control	44.0 (21.5-54.0)	44.23 (23.49)[j]		35.0 (25.25-56.0)	38.68 (19.58)[g]		30.0 (25.0-51.0)	36.96 (20.15)	
Intervention	47.0 (33.5-60.0)	48.22 (18.74)[k]		53.0 (38.0-66.0)	52.68 (18.27)[l]		50.0 (31.5-75.0)	51.05 (25.73)[l]	

[a]At admission to the program (T1), at the end of the intervention at 6 to 7 weeks (T2), and at the 3-month follow-up (T3).

[b]CPAQ: Chronic Pain Acceptance Questionnaire.

[c]n=53.

[d]n=28.

[e]n=27.

[f]n=56.

[g]n=22.

[h]n=23.

[i]EQ-VAS: EuroQol Visual Analogue Scale.

[j]n=47.

[k]n=50.

[l]n=19.

XSL•FO
RenderX

Table 6. Within-group differences and variations over time for the secondary outcomes.

Secondary outcome measure (scale)	T2–T1[a]			T3–T2[a]			T3–T1[a]		
	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value	Median (IQR)	Mean (SD)	*P* value
CPAQ[b]			.001			.14			.30
Control	−0.16 (−0.27 to 0.1)	0.05 (0.9)[c]		0.02 (−0.12 to 0.25)	0.08 (0.28)[d]		−0.08 (−0.24 to 0.09)	0.12 (0.92)[e]	
Intervention	0.22 (0.05 to 0.39)	0.38 (0.86)[f]		−0.1 (−0.24 to 0.16)	−0.22 (0.95)[g]		0.0 (−0.18 to 0.25)	−0.1 (0.86)[f]	
EQ-5D			.002			.17			.48
Control	−0.12 (−0.53 to 0.08)	−0.1 (0.69)[c]		0.0 (−0.03 to 0.22)	0.04 (0.34)[d]		0.0 (−0.4 to 0.45)	0.12 (0.69)[e]	
Intervention	0.31 (0.0 to 0.92)	0.43 (0.66)[f]		0.0 (−0.31 to 0.13)	−0.09 (0.61)[g]		0.0 (−0.1 to 0.35)	0.08 (0.75)[f]	
EQ-VAS[h]			.13			.49			.34
Control	0.14 (−0.77 to 0.49)	−0.27 (1.0)		−0.05 (−0.28 to 0.17)	−0.08 (0.31)		0.18 (−0.34 to 0.83)	0.18 (1.22)[i]	
Intervention	0.28 (−0.12 to 0.42)	0.25 (0.52)		−0.14 (−0.37 to 0.24)	−0.05 (0.4)		0.33 (−0.26 to 0.95)	0.18 (0.93)[d]	

[a]At admission to the program (T1), at the end of the intervention at 6 weeks (T2), and at the 3-month follow-up (T3).

[b]CPAQ: Chronic Pain Acceptance Questionnaire.

[c]n=26.

[d]n=16.

[e]n=27.

[f]n=21.

[g]n=17.

[h]EQ-VAS: EuroQol Visual Analogue Scale.

[i]n=24.

Between-Group Effects

With regard to pain acceptance (CPAQ), no significant differences were found between the 2 groups after treatment (68.23 vs 63.82; *P*=.14) or at 3 months following the intervention (64.48 vs 65.67; *P*=.47).

In terms of quality of life (EQ-5D), the intervention group showed significant improvement at the end of treatment (0.55 vs 0.41; *P*=.008), although these differences were not maintained after the 3-month follow-up (0.43 vs 0.39; *P*=.30).

The assessment of overall health state (EQ-VAS) registered in the daily records showed significant improvements in the intervention group compared with the control group (52.68 vs 38.68; *P*=.02) at the end of treatment, and these differences were maintained over time (51.05 vs 39.96; *P*=.02; Table 5).

Within-Group Effects

Regarding variations between the different phases, a positive effect was observed immediately following the intervention (T2–T1) in both acceptance (0.38 vs 0.05; *P*=.001) and quality of life (0.43 vs −0.01; *P*=.002), but the positive effect on overall satisfaction with health was not maintained (0.25 vs −0.27; *P*=.13).

No significant differences were found during follow-up for CPAQ (T3–T2: −0.22 vs 0.08 and *P*=.14; T3–T1: −0.1 vs 0.12 and *P*=.30), EQ-5D (T3–T2: −0.09 vs 0.04 and *P*=.17; T3–T1: 0.08 vs 0.12 and *P*=.48), or overall health state (T3–T2: −0.05 vs −0.08 and *P*=.49; T3–T1: 0.18 vs 0.18 and *P*=.34; Table 6).

Table 7 shows the Fisher-Freeman-Halton descriptive analysis of the proportion of participants with clinically significant improvement immediately after treatment according to the EQ-5D subscale. In the intervention group, clinical improvement in pain intensity ranged from 37% in moderate pain to 22.7% in severe pain (*P*=.04). The exact test results showed that a significantly higher proportion of participants who received the multimodal treatment improved in the mobility subscale (*P*=.04) and activities of daily living such as going to work, leisure time, and family activities (*P*=.045) immediately after receiving treatment. These improvements were not significant 3 months after the intervention.

XSL•FO
RenderX

Table 7. Proportion of patients to the EQ-5D subscale per time and group at admission (T1), immediately after treatment (T2), and 3 months after the intervention period (T3).

EQ-5D secondary outcome measure	T1			T2			T3		
	Intervention, n (%)	Control, n (%)	Overall *P* value	Intervention, n (%)	Control, n (%)	Overall *P* value	Intervention, n (%)	Control, n (%)	Overall *P* value
Pain			.34			.004			.08
Moderate pain (VAS[a]=4-7)	34 (61)	27 (51)		16 (37)	10 (78)		13 (34)	9 (53)	
A lot of pain (VAS≥8)	22 (39)	26 (49)		2 (23)	17 (64)		10 (46)	17 (67)	
Anxiety and depression			.90			.27			.12
I am not anxious or depressed	13 (23)	13 (25)		7 (32)	8 (29)		3 (13)	8 (30)	
I am moderately anxious or depressed	32 (57)	30 (58)		13 (59)	12 (43)		16 (67)	14 (52)	
I am very anxious or depressed	11 (20)	9 (17)		1 (9)	7 (29)		4 (21)	4 (19)	
Mobility			.14			.04			.07
I have no problem walking	9 (41)	10 (36)		12 (21)	13 (25)		7 (29)	8 (30)	
I have some trouble walking	43 (77)	39 (74)		13 (59)	18 (64)		16 (67)	19 (70)	
I have to be in bed	1 (2)	1 (2)		N/A[b]	N/A		1 (4)	N/A	
Personal care			.92			.77			.63
I have no problems with self-care	33 (60)	31 (59)		14 (65)	15 (56)		12 (50)	13 (50)	
I have some problems washing or dressing	22 (40)	21 (40)		8 (36)	12 (44)		12 (50)	14 (50)	
I am unable to wash or dress	N/A	1 (2)		N/A	N/A		N/A	N/A	
Everyday activities (work, study, household chores, free time, and family activities)			.14			.045			.86
I have no problem performing my daily activities	5 (9)	9 (17)		5 (24)	4 (14)		2 (8)	1 (4)	
I have some trouble performing my daily activities	49 (88)	38 (72)		16 (76)	20 (71)		20 (83)	23 (85)	
I am unable to perform my daily activities	2 (4)	6 (11)		N/A	4 (14)		2 (8)	3 (11)	

[a]VAS: Visual Analogue Scale.

[b]N/A: not applicable.

Discussion

Principal Findings

We performed a small randomized controlled clinical trial with a sample of mostly women (205/297, 69%). Randomization of both the sociodemographic characteristics of the study population and the study variables was homogeneous.

The results of the study clearly show that the main variable, catastrophizing, had a significant positive effect on the intervention group compared with the control group after the intervention and at the 3-month follow-up. This could be due to changes in cognition or maladaptive behavior that modified erroneous beliefs and decreased catastrophic thoughts. It should be noted that previous research has considered the beneficial effect of multimodal treatments in decreasing catastrophizing and fear. Thus, these results are in accordance with the more pronounced effect found in the intervention group [55].

These results are also in line with the findings of studies on the effects of monotherapies using ACT [56], mindfulness [57], physical activity, pharmacological therapy, and the health asset approach. Analysis of the results showed that the intervention group improved in catastrophizing, psychological flexibility, movement avoidance, pain interference in daily life, pain intensity, and quality of life [31,58-62].

The analysis showed that the intervention group improved in both perceived quality of life and pain acceptance after treatment. This finding indicates that patients who accept pain better exhibit greater psychological flexibility (ie, less activity avoidance), less psychological distress, and less disability. Moreover, we found that patients with higher pain acceptance (psychological flexibility; *P*=.001) reported a better quality of life (*P*=.002). This is in agreement with other studies that have shown that pain acceptance is a good predictor of a better quality of life [20,21,61].

XSL•FO
RenderX

However, the results were inconclusive for these last 2 variables. Specifically, although the effect sizes of quality of life and pain acceptance were significant from before treatment to after treatment in the intervention group, no changes were observed 3 months after the treatment. This lack of significance may be explained by the fact that perceived quality of life is a multidimensional phenomenon, and some indicators may therefore have had a greater influence on this variable, such as the low educational level and low annual income of the sample. These results are consistent with those of other studies suggesting that low educational level and low income are associated with poorer perceived health [63].

With regard to the main aim of this work, we hypothesized that the participants assigned to the intervention group would experience fewer catastrophizing thoughts and emotional distress and more acceptance of pain and improve their ability to perform activities of daily living according to self-management values. The results are encouraging as the effects of the treatment were largely maintained over time and reduced catastrophizing and the three dimensions of the PCS (helplessness, rumination, and magnification). This may suggest that a combination of these interventions promotes skills that result in behavior changes, at least in the medium term.

A priori, this relationship could be explained by the additive effect of the combined components of psychoeducational therapy, exercise, pharmacological treatment, and health assets. It may also be explained by the fear avoidance model in that exercise contributes to the reconceptualization of pain and reduces catastrophic thinking and the threat value of pain related to functional limitations [55]. Similarly, exercise may help divert attention away from rumination because of its attentional demands and mood effects, whereas the use of exercise as a self-management tool could increase self-efficacy and thereby reduce feelings of helplessness.

Physical activity likely helped the participants in our study learn about activities such as daily walking and low-intensity stretching to improve their physical, mental, and emotional well-being as well as acquire new habits and find alternatives to improve their quality of life [64]. These results are in line with recent studies and meta-analyses that show that intermittent or regular sessions of therapeutic exercise can reduce pain perception and sensitivity [65].

ACT and mindfulness were also a treatment goal in our study [60]. In the intervention group, both therapies had a positive effect on pain acceptance (psychological flexibility), improved health perception, and decreased levels of catastrophizing immediately after the intervention. This finding is in line with the available evidence suggesting that face-to-face or technology-based ACT is an effective self-management intervention for chronic pain and that it may be effective for the treatment of chronic pain [23,34,66].

Pain catastrophizing has been identified as a psychosocial factor that predicts adaptation to chronic pain and may contribute to its development and chronicity. In this regard, several studies [28,67] have examined the associations between pain catastrophizing and patient functioning and suggested that genetic and interpersonal factors, family history, pain duration, and comorbidities moderate pain and are likely to influence the strength of the association of the effects of catastrophizing on pain.

However, another explanation for the effect of our multimodal treatment may be related to patient-treatment matching. Specifically, broader-spectrum multimodal treatments have a greater likelihood of matching at least one treatment component to a patient's strength or deficit [68].

The overall health state self-reported by the intervention group improved after the intervention, although this effect was not maintained over time. This may be explained by the fact that the participants perceived an improvement in their health state when performing the proposed weekly activities during the treatment period because of the effort and time invested, which is known as the Hawthorne phenomenon [69]. Motivation, or the effect of feeling observed and cared for, may also have played a role. It is important to remember that it takes time to modify lifestyles and habits and see the benefits of change.

Moreover, the descriptive analysis of the EQ-5D dimensions showed improvement in at least three of the five core outcome domains (pain intensity, mobility, and activities of daily living) compared with the control group. This finding corroborates the influence of attitudes and beliefs that may affect the development of passive coping mechanisms such as rest and medication versus the ability to adopt active strategies such as physical activity and pain acceptance [70].

Strengths, Limitations, and Future Directions

This study has several strengths. First, we developed a multimodal program involving a variety of therapeutic activities that could be standardized and used in the future for other patients with chronic pain. An encouraging finding in this line was that the benefits of the treatment were largely maintained at follow-up, which may suggest that these interventions lead to the acquisition of skills that result in behavior change, at least in the medium term. Second, ICTs were used in combination with pharmacological and nonpharmacological therapeutic treatments in the outpatient setting to evaluate their impact on pain catastrophizing, pain acceptance, and quality of life. Third, the fact that the patients did the programmed activities in their natural environment is likely to promote self-efficacy, thus supporting the importance of the self-management component in interventions of this type [25].

One of the most important limitations of our study was the low response rate of the self-reported questionnaires sent via email. It should also be noted that the sample size was small. This may have affected the outcomes of the intervention and could explain the significant differences found between the groups. However, we cannot rule out the beneficial effect of the treatment on catastrophizing in the intervention group as a good prognosis of the disease and that, in the long term, these interventions (physical activity, psychoeducational therapies, pharmacological therapies, and health assets) are part of multi-professional treatments to achieve the desired effects. Nevertheless, as we used a multimodal therapy, we cannot really know the effect of each individual intervention on the outcomes, and more studies will be needed to determine the effects of specific interventions

XSL•FO
RenderX

in patients with chronic pain. Therefore, caution must be exercised when extrapolating our findings to the general population.

With regard to treatment adherence [5,71], we cannot reach reliable conclusions. Although the electronic records of the participants' access to and completion of the activities were not analyzed, we do know that 85% (83/98) of the participants completed the intervention program. This loss of participants may have occurred because the self-reported assessment was administered via email. Therefore, our treatment design could be improved by administering the questionnaire through the same mobile app, in interviews with the attending nurses, or by telephone follow-up.

According to our results and the available evidence, pharmacological treatments are most effective when they are part of an overall multidisciplinary pain management plan that also incorporates psychological, physical, and preventive components [8,72-74]. Clearly, patients with chronic pain should be informed and educated to enable them to make decisions about the most effective evidence-based strategy and ensure that their pain is treated and managed in the best possible manner.

A large number of studies have been conducted on ICT-based interventions to promote self-management in people with chronic pain [75-77], and there is evidence of high ICT acceptance and satisfaction [72,73]. Indeed, as technologies can assist and support people in their daily lives and at any time of the day [32], smartphones have become a very effective health care tool. In our smartphone app, we have selected evidence-based activities to address the various dimensions of chronic pain. The activities are easily reproducible in many environments and health care fields and can serve as complementary therapies for the comprehensive treatment of people with persistent pain. Nonetheless, it should be noted that self-management of pain is only effective when implemented from a multidisciplinary approach as treatment response is individual and there is no single approach that is beneficial for all patients with chronic pain. Therefore, for smartphone-based apps to be successful in promoting the self-management of chronic pain, we believe they should include self-monitoring, goal setting, skill training, social support, and educational components [78]. Moreover, many of these apps appear to have been developed without the involvement of patients and health care professionals and, to the best of our knowledge, few have been tested in randomized trials to evaluate their impact on health.

In future research, more attention should also be paid to the participants' gender as this could have affected our findings. It is well-known that gender is strongly related to access to care and treatment response and that, although many patients who experience pain are women, many stigmas are associated with pain in men. Therefore, it is important that we gain a better understanding of the role of gender in health care access as well as gender biases in diagnoses, patient-professional interactions, and treatment.

Further lines of research could improve the efficacy of multimodal chronic pain interventions based on new technologies, such as the refinement of treatments, the identification of moderating factors that might influence psychosocial variables, and their association with treatment adherence. To evaluate which groups of patients are more competent to self-manage a technology-based multimodal intervention would have been ideal. In the same vein, it is also worth mentioning that, in our study, we did not specifically assess satisfaction in relation to the use of technologies, and this could be a promising line for future research.

In future research, it might be interesting to analyze data not included in this study, such as mobile sensor data using accelerometers, gyrometers, and other sensors to monitor engagement and assess the timing of the activities.

Conclusions and Implications for Practice

The results of our study suggest that multidimensional treatments for adults with chronic pain improve catastrophizing, quality of life, and psychological flexibility immediately after treatment and that the effects are maintained for the primary outcome of catastrophizing for at least 3 months following treatment. This study has also shown that nonpharmacological treatments that include physical and psychoeducational therapy to promote active participation work well in combination with pharmacological strategies and that such interventions improve self-reliance in patients with chronic pain and help them cope constructively with pain.

The NO+Dolor app we have developed uses gamification to teach patients distraction methods and divert their attention away from pain as well as mindfulness techniques to improve pain acceptance. It also provides patients with a well-paced program of exercises and information on the proper use of medications to avoid side effects and helps them identify health assets to engage in pleasurable activities or find the resources they need. Moreover, the app-based mobile interventions we have presented here are flexible and self-directed, promote self-management in patients with chronic pain, and can be used to complement face-to-face pain treatments.

Preventive interventions for people with chronic pain designed from a salutogenic approach, are essential to promote well-being and prevent further decline in health throughout life.

Acknowledgments

The authors would like to thank all the patients who selflessly collaborated in this study and the professionals who participated in the creation of the audiovisual materials as this work would not have been possible without them. This study was funded by the Regional Ministry of Health by Resolution of May 18, 2017, of the General Secretariat for Biomedical Research, Development, and Innovation and in Health Sciences of Andalusia with the collaboration of the Biomedical Research Foundation of Cordoba.

XSL•FO
RenderX

The study has the authorization of the corresponding Research Ethics Committee and complies with the current law on data protection (promoter: Biomedical Research Foundation of Cordoba; protocol code: PI-0447-2017; study code: S0452). This study was conducted according to the guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of the Foundation for Biomedical Research of Cordoba, Andalusia, Spain (CIF G-14825277; protocol code PI-0447-2017, resolution December 22, 2017, and published in the official bulletin of the Junta de Andalusia, Spain, October 5, 2018). Informed consent was obtained from all participants involved in the study. Written informed consent was obtained from the patients to publish this paper.

Data Availability

The data supporting the reported results can be found in the electronic notebook of the Data Entry Manager of the Maimonides Biomedical Research Institute of Cordoba. The statistician, principal investigator, and collaborators who evaluated the results did not participate in patient recruitment.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key contents of the intervention.
[PDF File (Adobe PDF File), 81 KB - jmir_v24i5e36114_app1.pdf]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 438 KB - jmir_v24i5e36114_app2.pdf]

References

1. Breivik H, Eisenberg E, O'Brien T, OPENMinds. The individual and societal burden of chronic pain in Europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. BMC Public Health 2013 Dec 24;13:1229 [FREE Full text] [doi: 10.1186/1471-2458-13-1229] [Medline: 24365383]
2. van Hecke O, Torrance N, Smith BH. Chronic pain epidemiology and its clinical relevance. Br J Anaesth 2013 Jul;111(1):13-18 [FREE Full text] [doi: 10.1093/bja/aet123] [Medline: 23794640]
3. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain 2020 Sep 01;161(9):1976-1982 [FREE Full text] [doi: 10.1097/j.pain.0000000000001939] [Medline: 32694387]
4. Nicholas M, Vlaeyen JW, Rief W, Barke A, Aziz Q, Benoliel R, IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic primary pain. Pain 2019 Jan;160(1):28-37. [doi: 10.1097/j.pain.0000000000001390] [Medline: 30586068]
5. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Jan 14;1(1):CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub2] [Medline: 28087891]
6. Mills SE, Nicolson KP, Smith BH. Chronic pain: a review of its epidemiology and associated factors in population-based studies. Br J Anaesth 2019 Aug;123(2):e273-e283 [FREE Full text] [doi: 10.1016/j.bja.2019.03.023] [Medline: 31079836]
7. Henschke N, Kamper SJ, Maher CG. The epidemiology and economic consequences of pain. Mayo Clin Proc 2015 Jan;90(1):139-147. [doi: 10.1016/j.mayocp.2014.09.010] [Medline: 25572198]
8. Gauthier K, Dulong C, Argáez C. Multidisciplinary treatment programs for patients with chronic non-malignant pain: a review of clinical effectiveness, cost-effectiveness, and guidelines – an update [Internet]. Ottawa, ON, Canada: Canadian Agency for Drugs and Technologies in Health; May 10, 2019.
9. Häuser W, Bernardy K, Arnold B, Offenbächer M, Schiltenwolf M. Efficacy of multicomponent treatment in fibromyalgia syndrome: a meta-analysis of randomized controlled clinical trials. Arthritis Rheum 2009 Feb 15;61(2):216-224 [FREE Full text] [doi: 10.1002/art.24276] [Medline: 19177530]
10. Wells C, Ford C. Strategies for the reduction or discontinuation of opioids: guidelines. Canada's Drug and Health Technology Agency. Ottawa: CADTH; 2017 Apr 5. URL: https://www.cadth.ca/sites/default/files/pdf/htis/2017/RB1078%20-%20Opioid%20Tapering%20Final.pdf [accessed 2022-04-10]
11. Ross EL, Jamison RN, Nicholls L, Perry BM, Nolen KD. Clinical integration of a smartphone app for patients with chronic pain: retrospective analysis of predictors of benefits and patient engagement between clinic visits. J Med Internet Res 2020 Apr 16;22(4):e16939 [FREE Full text] [doi: 10.2196/16939] [Medline: 32297871]
12. Turk DC, Fillingim RB, Ohrbach R, Patel KV. Assessment of psychosocial and functional impact of chronic pain. J Pain 2016 Sep;17(9 Suppl):T21-T49. [doi: 10.1016/j.jpain.2016.02.006] [Medline: 27586830]

XSL•FO
RenderX

13. Chronic pain in Canada: laying a foundation for action? - A report by the Canadian Pain Task Force, June 2019. Health Canada. 2019. URL: https://tinyurl.com/uae72c3s [accessed 2022-04-10]
14. Deckert S, Kaiser U, Kopkow C, Trautmann F, Sabatowski R, Schmitt J. A systematic review of the outcomes reported in multimodal pain therapy for chronic pain. Eur J Pain 2016 Jan;20(1):51-63. [doi: 10.1002/ejp.721] [Medline: 26031689]
15. Management of chronic pain. Scottish Intercollegiate Guidelines Network (SIGN). 2013 Dec. URL: https://www.sign.ac.uk/our-guidelines/management-of-chronic-pain/ [accessed 2022-04-10]
16. Booth J, Moseley GL, Schiltenwolf M, Cashin A, Davies M, Hübscher M. Exercise for chronic musculoskeletal pain: a biopsychosocial approach. Musculoskeletal Care 2017 Dec;15(4):413-421. [doi: 10.1002/msc.1191] [Medline: 28371175]
17. Hearing CM, Chang WC, Szuhany KL, Deckersbach T, Nierenberg AA, Sylvia LG. Physical exercise for treatment of mood disorders: a critical review. Curr Behav Neurosci Rep 2016 Dec;3(4):350-359 [FREE Full text] [doi: 10.1007/s40473-016-0089-y] [Medline: 28503402]
18. Mackey S. National Pain Strategy Task Force: the strategic plan for the IOM Pain Report. Pain Med 2014 Jul;15(7):1070-1071. [doi: 10.1111/pme.12490] [Medline: 25059928]
19. Mills S, Torrance N, Smith BH. Identification and management of chronic pain in primary care: a review. Curr Psychiatry Rep 2016 Feb;18(2):22 [FREE Full text] [doi: 10.1007/s11920-015-0659-9] [Medline: 26820898]
20. Vowles KE, Fink BC, Cohen LL. Acceptance and Commitment Therapy for chronic pain: a diary study of treatment process in relation to reliable change in disability. J Contextual Behav Sci 2014 Apr;3(2):74-80 [FREE Full text] [doi: 10.1016/j.jcbs.2014.04.003] [Medline: 27818931]
21. McCracken LM, Vowles KE, Eccleston C. Acceptance of chronic pain: component analysis and a revised assessment method. Pain 2004 Jan;107(1-2):159-166. [doi: 10.1016/j.pain.2003.10.012] [Medline: 14715402]
22. Majeed MH, Ali AA, Sudak DM. Mindfulness-based interventions for chronic pain: evidence and applications. Asian J Psychiatr 2018 Feb;32:79-83. [doi: 10.1016/j.ajp.2017.11.025] [Medline: 29220782]
23. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
24. Adler-Neal AL, Waugh CE, Garland EL, Shaltout HA, Diz DI, Zeidan F. The role of heart rate variability in mindfulness-based pain relief. J Pain 2020;21(3-4):306-323 [FREE Full text] [doi: 10.1016/j.jpain.2019.07.003] [Medline: 31377215]
25. Moore SM, Schiffman R, Waldrop-Valverde D, Redeker NS, McCloskey DJ, Kim MT, et al. Recommendations of common data elements to advance the science of self-management of chronic conditions. J Nurs Scholarsh 2016 Sep;48(5):437-447 [FREE Full text] [doi: 10.1111/jnu.12233] [Medline: 27486851]
26. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
27. Hadi MA, McHugh GA, Closs SJ. Impact of chronic pain on patients' quality of life: a comparative mixed-methods study. J Patient Exp 2019 Jun;6(2):133-141 [FREE Full text] [doi: 10.1177/2374373518786013] [Medline: 31218259]
28. Martinez-Calderon J, Jensen MP, Morales-Asencio JM, Luque-Suarez A. Pain catastrophizing and function in individuals with chronic musculoskeletal pain: a systematic review and meta-analysis. Clin J Pain 2019 Mar;35(3):279-293. [doi: 10.1097/AJP.0000000000000676] [Medline: 30664551]
29. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
30. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. In: Database of Abstracts of Reviews of Effects (DARE): Quality-assessed Reviews. York, UK: Centre for Reviews and Dissemination; 1995.
31. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain. part 2: 11-month follow-up results of a randomized trial. J Med Internet Res 2013 Mar 28;15(3):e72 [FREE Full text] [doi: 10.2196/jmir.2442] [Medline: 23538392]
32. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
33. Luxton DD, McCann RA, Bush NE, Mishkind MC, Reger GM. mHealth for mental health: integrating smartphone technology in behavioral healthcare. Prof Psychol Res Pract 2011 Dec;42(6):505-512 [FREE Full text] [doi: 10.1037/a0024485]
34. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: a systematic review. Internet Interv 2016 May;4:17-34 [FREE Full text] [doi: 10.1016/j.invent.2015.12.001] [Medline: 30135787]
35. Slattery BW, Haugh S, O'Connor L, Francis K, Dwyer CP, O'Higgins S, et al. An evaluation of the effectiveness of the modalities used to deliver electronic health interventions for chronic pain: systematic review with network meta-analysis. J Med Internet Res 2019 Jul 17;21(7):e11086 [FREE Full text] [doi: 10.2196/11086] [Medline: 31317869]

XSL•FO
RenderX

36. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008 Feb;9(2):105-121. [doi: 10.1016/j.jpain.2007.09.005] [Medline: 18055266]
37. Kristjánsdóttir Ó, Fors E, Eide E, Finset A, van Dulmen S, Wigers SH, et al. Written online situational feedback via mobile phone to support self-management of chronic widespread pain: a usability study of a web-based intervention. BMC Musculoskelet Disord 2011 Feb 25;12:51 [FREE Full text] [doi: 10.1186/1471-2474-12-51] [Medline: 21352516]
38. Morcillo-Muñoz Y, Castellano MH, Exposito FJ, Sanchez-Guarnido AJ, Alcantara MG, Baena-Parejo MI. Multimodal interventions to improve the management of chronic non-malignant pain in primary care using participatory research. Clin Pract 2021 Aug 26;11(3):561-581 [FREE Full text] [doi: 10.3390/clinpract11030072] [Medline: 34449567]
39. Dorado K, Schreiber KL, Koulouris A, Edwards RR, Napadow V, Lazaridou A. Interactive effects of pain catastrophizing and mindfulness on pain intensity in women with fibromyalgia. Health Psychol Open 2018 Oct 22;5(2):2055102918807406 [FREE Full text] [doi: 10.1177/2055102918807406] [Medline: 30364853]
40. Trompetter HR, Bohlmeijer ET, Fox JP, Schreurs KM. Behav Res Ther 2015 Nov;74:50-59. [doi: 10.1016/j.brat.2015.09.001] [Medline: 26409158]
41. American Society of Anesthesiologists Task Force on Chronic Pain Management. Practice guidelines for chronic pain management. A report by the American Society of Anesthesiologists Task Force on Pain Management, Chronic Pain Section. Anesthesiology 1997 Apr;86(4):995-1004. [doi: 10.1097/ALN.0b013e3181c43103] [Medline: 9105246]
42. Busse JW, Craigie S, Juurlink DN, Buckley DN, Wang L, Couban RJ, et al. Guideline for opioid therapy and chronic noncancer pain. CMAJ 2017 May 08;189(18):E659-E666 [FREE Full text] [doi: 10.1503/cmaj.170363] [Medline: 28483845]
43. Clinical best practice guidelines: assessment and management of pain. Registered Nurses' Association of Ontario (RNAO). 2013. URL: https://rnao.ca/bpg/guidelines/assessment-and-management-pain [accessed 2022-04-10]
44. Blickem C, Dawson S, Kirk S, Vassilev I, Mathieson A, Harrison R, et al. What is asset-based community development and how might it improve the health of people with long-term conditions? A realist synthesis. SAGE Open 2018 Jul 06;8(3):215824401878722 [FREE Full text] [doi: 10.1177/2158244018787223]
45. Mittelmark MB, Bull T, Bouwman L. Emerging ideas relevant to the Salutogenic model of health. In: Mittelmark MB, Sagy S, Eriksson M, Bauer GF, Pelikan JM, Lindström B, et al, editors. The handbook of Salutogenesis. Cham, Switzerland: Springer; Sep 3, 2017.
46. Biomedical Research Foundation of Cordoba. 2019. URL: https://apps.apple.com/es/app/no-más-dolor-edcno/id1458096026 [accessed 2019-05-17]
47. García Campayo J, Rodero B, Alda M, Sobradiel N, Montero J, Moreno S. [Validation of the Spanish version of the Pain Catastrophizing Scale in fibromyalgia]. Med Clin (Barc) 2008 Oct 18;131(13):487-492. [doi: 10.1157/13127277] [Medline: 19007576]
48. Lami MJ, Martinez MP, Miró E, Sánchez AI. Spanish version of the Pain Catastrophizing Scale: psychometric study in healthy women. Behav Psychol 2013 Apr 1;22(1):137-156.
49. Sullivan MJ, Bishop SR, Pivik J. The Pain Catastrophizing Scale: development and validation. Psychol Assess 1995;7(4):524-532 [FREE Full text] [doi: 10.1037/1040-3590.7.4.524]
50. Arteta J, Cobos B, Hu Y, Jordan K, Howard K. Evaluation of how depression and anxiety mediate the relationship between pain catastrophizing and prescription opioid misuse in a chronic pain population. Pain Med 2016 Feb;17(2):295-303. [doi: 10.1111/pme.12886] [Medline: 26235471]
51. Barraca J. Spanish adaptation of the acceptance and action questionnaire (AAQ). Int J Psychol Psychol Ther 2004;4(3):505-515.
52. Badia X, Roset M, Montserrat S, Herdman M, Segura A. [The Spanish version of EuroQol: a description and its applications. European Quality of Life scale]. Med Clin (Barc) 1999;112 Suppl 1:79-85. [Medline: 10618804]
53. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
54. Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics 1982 Dec;38(4):963-974. [Medline: 7168798]
55. Watson JA, Ryan CG, Cooper L, Ellington D, Whittle R, Lavender M, et al. Pain neuroscience education for adults with chronic musculoskeletal pain: a mixed-methods systematic review and meta-analysis. J Pain 2019 Oct;20(10):1140.e1-1140.e2. [doi: 10.1016/j.jpain.2019.02.011] [Medline: 30831273]
56. Bhatia A, Kara J, Janmohamed T, Prabhu A, Lebovic G, Katz J, et al. User engagement and clinical impact of the manage my pain app in patients with chronic pain: a real-world, multi-site trial. JMIR Mhealth Uhealth 2021 Mar 04;9(3):e26528 [FREE Full text] [doi: 10.2196/26528] [Medline: 33661130]
57. Winter N, Russell L, Ugalde A, White V, Livingston P. Engagement strategies to improve adherence and retention in web-based mindfulness programs: systematic review. J Med Internet Res 2022 Jan 12;24(1):e30026 [FREE Full text] [doi: 10.2196/30026] [Medline: 35019851]
58. Khoo EL, Small R, Cheng W, Hatchard T, Glynn B, Rice DB, et al. Comparative evaluation of group-based mindfulness-based stress reduction and cognitive behavioural therapy for the treatment and management of chronic pain: a systematic review

XSL•FO
RenderX

and network meta-analysis. Evid Based Ment Health 2019 Feb;22(1):26-35. [doi: 10.1136/ebmental-2018-300062] [Medline: 30705039]
59. Oerlemans S, van Cranenburgh O, Herremans PJ, Spreeuwenberg P, van Dulmen S. Intervening on cognitions and behavior in irritable bowel syndrome: a feasibility trial using PDAs. J Psychosom Res 2011 Mar;70(3):267-277. [doi: 10.1016/j.jpsychores.2010.09.018] [Medline: 21334498]
60. Schütze R, Rees C, Smith A, Slater H, Campbell JM, O'Sullivan P. How can we best reduce pain catastrophizing in adults with chronic noncancer pain? A systematic review and meta-analysis. J Pain 2018 Mar;19(3):233-256. [doi: 10.1016/j.jpain.2017.09.010] [Medline: 29122652]
61. Bartley EJ, Palit S, Fillingim RB, Robinson ME. Multisystem resiliency as a predictor of physical and psychological functioning in older adults with chronic low back pain. Front Psychol 2019 Aug 22;10:1932 [FREE Full text] [doi: 10.3389/fpsyg.2019.01932] [Medline: 31507491]
62. Aghayousefi A, Tarkhan M, Mohammadi N, Afshar H. The role of psychological inflexibility and pain acceptance in predicting of resiliency in chronic pain patients. Q J Health Psychol 2016 Jul;5(17):23-38.
63. Aguilar-Palacio I, Carrera-Lasfuentes P, Rabanaque MJ. [Self-rated health and educational level in Spain: trends by autonomous communities and gender (2001-2012)]. Gac Sanit 2015;29(1):37-43 [FREE Full text] [doi: 10.1016/j.gaceta.2014.07.004] [Medline: 25127554]
64. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee IM, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
65. Wewege MA, Jones MD. Exercise-induced hypoalgesia in healthy individuals and people with chronic musculoskeletal pain: a systematic review and meta-analysis. J Pain 2021 Jan;22(1):21-31. [doi: 10.1016/j.jpain.2020.04.003] [Medline: 32599154]
66. McCracken LM, Sato A, Taylor GJ. A trial of a brief group-based form of acceptance and commitment therapy (ACT) for chronic pain in general practice: pilot outcome and process results. J Pain 2013 Nov;14(11):1398-1406 [FREE Full text] [doi: 10.1016/j.jpain.2013.06.011] [Medline: 24035351]
67. Martinez-Calderon J, Meeus M, Struyf F, Miguel Morales-Asencio J, Gijon-Nogueron G, Luque-Suarez A. The role of psychological factors in the perpetuation of pain intensity and disability in people with chronic shoulder pain: a systematic review. BMJ Open 2018 Apr 13;8(4):e020703 [FREE Full text] [doi: 10.1136/bmjopen-2017-020703] [Medline: 29654040]
68. Jeffery MM, Butler M, Stark A, Kane RL. Multidisciplinary pain programs for chronic noncancer pain. Rockville, MD, USA: Agency for Healthcare Research and Quality; Sep 2011.
69. Luciano JV, Martínez N, Peñarrubia-María MT, Fernández-Vergel R, García-Campayo J, Verduras C, FibroQoL Study Group. Effectiveness of a psychoeducational treatment program implemented in general practice for fibromyalgia patients: a randomized controlled trial. Clin J Pain 2011 Jun;27(5):383-391. [doi: 10.1097/AJP.0b013e31820b131c] [Medline: 21317775]
70. Blyth FM, March LM, Nicholas MK, Cousins MJ. Self-management of chronic pain: a population-based study. Pain 2005 Feb;113(3):285-292. [doi: 10.1016/j.pain.2004.12.004] [Medline: 15661435]
71. Aitken D, Buchbinder R, Jones G, Winzenberg T. Interventions to improve adherence to exercise for chronic musculoskeletal pain in adults. Aust Fam Physician 2015;44(1-2):39-42 [FREE Full text] [Medline: 25688958]
72. Gatchel RJ, McGeary DD, McGeary CA, Lippe B. Interdisciplinary chronic pain management: past, present, and future. Am Psychol 2014;69(2):119-130. [doi: 10.1037/a0035514] [Medline: 24547798]
73. Access to and availability of non-pharmacological treatments for chronic non-cancer pain in Canada: an environmental scan - Health Technology Review. Canada's Drug and Health Technology Agency. 2021 Nov. URL: https://cadth.ca/access-and-availability-non-pharmacological-treatments-chronic-non-cancer-pain-canada-0 [accessed 2022-04-10]
74. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Global Health, Board on Health Sciences Policy, Global Forum on Innovation in Health Professional Education, Forum on Neuroscience and Nervous System Disorders. In: Stroud C, Posey Norris SM, Bain L, editors. The role of nonpharmacological approaches to pain management: proceedings of a workshop. Washington, DC, USA: National Academies Press; Apr 12, 2019.
75. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
76. Totten AM, Hansen RN, Wagner J, Stillman L, Ivlev I, Davis-O'Reilly C, et al. Telehealth for acute and chronic care consultations. Rockville, MD, USA: Agency for Healthcare Research and Quality; Apr 2019.
77. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Mirkovic J, Børøsund E, et al. Patients' needs and requirements for eHealth pain management interventions: qualitative study. J Med Internet Res 2019 Apr 01;21(4):e13205 [FREE Full text] [doi: 10.2196/13205] [Medline: 30877780]
78. 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] [Medline: 22209829]

XSL•FO
RenderX

Abbreviations

ACT: acceptance and commitment therapy
CPAQ: Chronic Pain Acceptance Questionnaire
EQ-VAS: EuroQol Visual Analogue Scale
ICT: information and communication technology
PCS: Pain Catastrophizing Scale
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 02.01.22; peer-reviewed by H Mehdizadeh, V Yadav, K Albakri, A Naser; comments to author 17.02.22; revised version received 06.03.22; accepted 31.03.22; published 11.05.22.

Please cite as:
Morcillo-Muñoz Y, Sánchez-Guarnido AJ, Calzón-Fernández S, Baena-Parejo I
Multimodal Chronic Pain Therapy for Adults via Smartphone: Randomized Controlled Clinical Trial
J Med Internet Res 2022;24(5):e36114
URL: https://www.jmir.org/2022/5/e36114
doi:10.2196/36114
PMID:35373

©Yolanda Morcillo-Muñoz, Antonio José Sánchez-Guarnido, Silvia Calzón-Fernández, Isabel Baena-Parejo. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 11.05.2022. 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.

XSL•FO
RenderX

Original Paper

Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation

Malte L Schmieding[1], MD; Marvin Kopka[1,2], BSc, MSc; Konrad Schmidt[3,4], MD; Sven Schulz-Niethammer[5], BSc; Felix Balzer[1*], MSc, MD, PhD; Markus A Feufel[5*], Dipl-Ing (FH), MSc, PhD

[1]Institute of Medical Informatics, Charité - Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[2]Cognitive Psychology and Ergonomics, Department of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

[3]Institute of General Practice and Family Medicine, Jena University Hospital, Germany, Jena, Germany

[4]Institute of General Practice and Family Medicine, Charité - Universitätsmedizin Berlin, Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[5]Division of Ergonomics, Department of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

[*]these authors contributed equally

Corresponding Author:
Malte L Schmieding, MD
Institute of Medical Informatics
Charité - Universitätsmedizin Berlin
Freie Universität Berlin and Humboldt-Universität zu Berlin
Charitéplatz 1
Berlin, 10117
Germany
Phone: 49 30450 570425
Email: malte.schmieding@charite.de

Abstract

Background: Symptom checkers are digital tools assisting laypersons in self-assessing the urgency and potential causes of their medical complaints. They are widely used but face concerns from both patients and health care professionals, especially regarding their accuracy. A 2015 landmark study substantiated these concerns using case vignettes to demonstrate that symptom checkers commonly err in their triage assessment.

Objective: This study aims to revisit the landmark index study to investigate whether and how symptom checkers' capabilities have evolved since 2015 and how they currently compare with laypersons' stand-alone triage appraisal.

Methods: In early 2020, we searched for smartphone and web-based applications providing triage advice. We evaluated these apps on the same 45 case vignettes as the index study. Using descriptive statistics, we compared our findings with those of the index study and with publicly available data on laypersons' triage capability.

Results: We retrieved 22 symptom checkers providing triage advice. The median triage accuracy in 2020 (55.8%, IQR 15.1%) was close to that in 2015 (59.1%, IQR 15.5%). The apps in 2020 were less risk averse (odds 1.11:1, the ratio of overtriage errors to undertriage errors) than those in 2015 (odds 2.82:1), missing >40% of emergencies. Few apps outperformed laypersons in either deciding whether emergency care was required or whether self-care was sufficient. No apps outperformed the laypersons on both decisions.

Conclusions: Triage performance of symptom checkers has, on average, not improved over the course of 5 years. It decreased in 2 use cases (advice on when emergency care is required and when no health care is needed for the moment). However, triage capability varies widely within the sample of symptom checkers. Whether it is beneficial to seek advice from symptom checkers depends on the app chosen and on the specific question to be answered. Future research should develop resources (eg, case vignette repositories) to audit the capabilities of symptom checkers continuously and independently and provide guidance on when and to whom they should be recommended.

(J Med Internet Res 2022;24(5):e31810) doi:10.2196/31810

KEYWORDS

digital health; triage; symptom checker; patient-centered care; eHealth apps; mobile phone

XSL•FO
RenderX

Introduction

Background

The use of patient-facing clinical decision support tools has become more and more prevalent in recent years. Tools assisting laypersons in their self-assessment on whether and where to seek professional medical care and for what diagnoses based on the users' input of symptoms and medical history are termed symptom checkers. Especially at the beginning of the COVID-19 pandemic, such tools were developed to assist patients in deciding whether to call emergency services with symptoms indicative of a COVID-19 infection or whether self-isolation is required [1-4]. Although the 2021 World Health Organization global report on artificial intelligence for health [5] mentions symptom checkers explicitly only in this narrow context of outbreak response, symptom checkers have been available (and researched [6]) for more than a decade and typically address a broader spectrum of diseases.

Symptom checkers are becoming increasingly popular, with approximately 13% of the German adult population already having consulted an app for self-diagnosis [7]. Such apps are commonly used as a means of self-information and guidance through the health care system [8], although other potential use cases such as syndromic surveillance have been described as well [9,10]. Notably, some integrated delivery networks (health care networks) in the United States have begun to incorporate symptom checkers as a service for their members, be it for educational purposes or to improve their members' experience of their patient journey, for example, by guidance on where and how urgently to seek care within the network on the symptom checker's suggestion [11-13]. A recent study among Finnish primary care leaders of institutions integrating symptom checkers into their services demonstrates their support for the use of these systems [14]. Despite their popularity, symptom checkers also face concerns from both patients and health care professionals [15-17]. Insufficient accuracy of the advice provided is a commonly raised concern.

Although no clear framework has yet been established on evaluating the accuracy of symptom checker apps' advice [18], a common first approach has been to test such systems on short fictitious patient descriptions (case vignettes), mirroring an approach to assess the reliability of diagnostic decision support systems for health care professionals [19,20], variability of initial diagnostic impressions [21], and diagnostic ability among physicians [22]. Independent studies using this approach suggest that advice from most symptom checker apps is rather unreliable, both for diagnosis and triage (ie, the assessment of urgency) [23-26]. A recent study with a slightly different approach, assessing 2 symptom checkers on information abstracted from medical records, came to the same conclusion [27]. Accordingly, 2 (systematic) reviews on the currently available evidence advise against using web-based triage systems in lieu of traditional urgency assessment means and emphasize the clinical risks that symptom checker use might pose [28,29].

Objective

A key source of these and other reviews [30,31] on symptom checkers is a study by Semigran et al [23] published in 2015. They found that symptom checkers were rather risk averse at the time and reported an aggregated triage accuracy of 57% and diagnostic accuracy of providing the correct diagnostic suggestion first at 34%. A recent study assessing 4 symptom checkers on ophthalmologic case vignettes in 2018 and 2020 suggests that the capabilities of symptom checkers have not improved during this time frame [25]. Although other relevant studies competitively comparing symptom checkers' accuracy have been published since 2015 [24,32,33], their chosen methodology (eg, sampling of apps and definition of triage levels) hinders a direct comparison with the data from the study by Semigran et al [23]. Thus, our study aims to revisit the landmark study by Semigran et al [23] to investigate whether and how symptom checkers' capabilities have evolved since 2015 and how they compare with human decision makers.

Methods

Data Collection on Symptom Checker Performance in 2020

Search and Selection Criteria for Symptom Checkers

Between February and March 2020, we systematically searched for symptom checker apps and websites capable of providing triage advice following the approach of Semigran et al [23]. To identify smartphone apps, we entered *symptom checker* and *medical diagnoses* as keywords in Google Play Store and the US, UK, and German versions of the Apple App Store and screened the first 240 results provided using the same cutoff as Semigran et al [23]. Symptom checker apps had to be freely available in English. We excluded apps that did not provide triage advice or only addressed specific complaints (eg, skin conditions) or specialties (eg, apps for orthopedics). Unlike Semigran et al [23], we also dismissed apps when the number of downloads was <100,000 or when the app had received unsatisfactory reviews (<4 stars), as both features indicate that the app might not be used commonly.

To identify web-based symptom checkers, we searched Google and Google Scholar with the same keywords and screened the first 300 results. Symptom checkers that were included in the study by Semigran et al [23] were searched by their name and included if still publicly available. Symptom checkers mentioned in other scientific studies or known to the authors but not found during the search strategy were included if they met the inclusion criteria as described earlier.

Modification of Clinical Vignettes

To analyze the diagnostic and triage accuracy of symptom checker apps, we used the same 45 short descriptions of fictitious patients and their complaints (*case vignettes*) that were used in the original study by Semigran et al [23]. As the interpretation of a vignette may change if it does not include a particular piece of information that is requested by a symptom checker (eg, the chief complaint's duration), we used several measures to augment the case description and, thus, improve

XSL•FO
RenderX

the comparability of vignettes across the different symptom checkers. First, 2 authors (KS and MS), both physicians, complemented indication-specific details that they anticipated to be relevant for triage decisions by advanced symptom checkers, such as the onset of the chief complaint or the severity of pain. In addition, each case vignette was assigned a chief complaint as some symptom checkers require this as input. When applicable, we adopted the chief complaint assigned by Hill et al [24], a 2020 study that also built upon the vignette sample from the study by Semigran et al [23]. However, we retained the gold standard solutions for the correct diagnosis and the 3-tiered triage-level definitions as defined in the index study [23].

Assessing Diagnostic and Triage Accuracy of Symptom Checker Apps in 2020

A research assistant with no clinical training entered the case vignettes into the symptom checker apps between June and December 2020. In advance, a set of rules was defined on how to handle ambiguity during data entry and outcome interpretation; for example, when symptom checkers requested information that was not provided by the vignette or an app's diagnostic suggestions were synonyms or umbrella terms for the gold standard diagnostic solution (Multimedia Appendix 1 [23,24,27,34-36]). Importantly, when a symptom checker app linked its triage advice to its diagnostic suggestions, we rated the triage advice for the first diagnosis it listed, assuming that this is the diagnosis that the app considers most likely and therefore the triage advice most relevant. This rule marks an exception from our general approach to retain the procedure of Semigran et al [23]. Semigran et al [23] used the most urgent triage level when the suggested diagnoses were linked with different triage suggestions. They argued that "in almost all of the cases the most urgent triage suggestion was listed first" [23]. However, we did not observe this and considered the triage advice linked to the diagnosis listed first. However, this divergence only affects 3 apps (*K Health*, *HealthTap*, and *WebMD*), as most apps provide triage advice not linked to diagnostic suggestions or also provide an overall triage appraisal.

As the apps use different classifications of triage urgency, we mapped all triage advice definitions of the assessed symptom checkers into the 3 categories that were defined by Semigran et al [23] (ie, emergency, nonemergency, and self-care). When symptom checkers provided triage advice that could not be matched to the 3 categories (eg, if a symptom checker identified emergency cases but could not specify whether self-care was sufficient or nonemergency care should be advised and when it deemed emergency care unnecessary), unspecified answers were counted as incorrect. To control for this decision, we conducted our main analyses twice, excluding and including such symptom checkers, and we report or provide both in the Multimedia Appendix 1.

Comparator Data Sets for Symptom Checker Performance

We compared our data on symptom checker performance in 2020 with three publicly available data sets: (1) Semigran et al [23] (the index study) evaluated the diagnostic and triage accuracy of 23 symptom checker apps in 2015 using 45 case vignettes; (2) Hill et al [24,34] evaluated 36 symptom checker apps in 2020 on 48 clinical vignettes, using some of the vignettes compiled by Semigran et al [23] and new case vignettes; and (3) Schmieding et al [35] evaluated laypersons' abilities to triage the same 45 case vignettes used by Semigran et al [23]. Although Semigran et al [23] used a 3-tiered classification of triage levels to set the gold standard solution (emergency care required, nonemergency care required, and self-care appropriate) and the study by Schmieding et al [36] retained this classification, the study by Hill et al [24] used a 4-tiered classification of triage levels (emergency, urgent, nonurgent, and self-care), thereby hindering a direct comparison of triage capability with the other data sets. Not all symptom checkers included in the first and second data sets and gave advice on both triage and diagnosis. In addition, in both studies, some symptom checkers never returned *self-care appropriate* as triage advice. A detailed description of these data sets can be found in the Multimedia Appendix 1. We made our data set publicly available via a web-based open data repository [37].

Data Analysis

Data were cleaned and explored with R (version 4.0.0) [38] and the *tidyverse* packages [39]. Figures were created using the package *ggplot2* [40].

Direct Comparison of Symptom Checker Triage Performance Between 2015 and 2020

We defined the triage accuracy of a symptom checker app as the proportion of vignettes to which a symptom checker app provided the correct triage advice in relation to all vignettes to which the app provided triage advice. In other words, vignettes that a symptom checker could not triage were omitted from the denominator. Given that our data from 2020 was based on the same 45 case vignettes and we retained the same definition of urgency levels and the gold standard solutions as set by Semigran et al [23], a direct comparison between these data sets was possible to assess the evolution of symptom checkers' triage capability between 2015 and 2020.

We calculated the median and IQRs of the apps' triage accuracies for both data sets. To assess whether symptom checker apps were still as risk averse as reported in 2015, we calculated the odds of overtriage to undertriage, where overtriage refers to advice to seek a more urgent level of care than necessary, and undertriage refers to advice to seek care with less urgency than appropriate. In addition, we created confusion tables outlining which triage advice was provided during the evaluations of case vignettes from each of the 3 triage levels. In both the analysis of risk aversion and the confusion table, we excluded apps that did not provide self-care advice in our results reported here. For example, we excluded the symptom checker *iTriage* in this analysis, as it always advised to seek emergency care in the data from the study by Semigran et al [23] and thus potentially masks an interesting trend observable in those apps capable of providing triage advice for all 3 triage levels. Analyses including all apps can be found in the Multimedia Appendix 1.

XSL•FO
RenderX

Comparison of Triage Accuracy for Binary Triage Decisions

A 3-tiered triage classification as used by Semigran et al [23] (emergency, nonemergency, and self-care) and retained for our data collection comes with 2 downsides. First, a direct comparison of results from studies with different classifications of urgency levels is hindered, and second, common metrics of signal detection theory (eg, sensitivity and specificity) cannot be calculated. To facilitate a comparison of triage accuracy across studies with different triage definitions (eg, emergency, urgent care, nonurgent care, and self-care, as used in the study by Hill et al [24]), we created 2 binary triage accuracy measures: whether a symptom checker (or layperson) can differentiate between cases requiring emergency care (decision 1) and between cases where self-care was sufficient or professional medical care should be sought (decision 2). These 2 measures represent common triage decisions users of symptom checkers face [8]. Accordingly, case vignettes with gold standard urgency levels of *urgent care*, *nonemergency care*, *nonurgent care*, and *self-care* are counted as not requiring emergency care, whereas all urgency levels except *self-care* were counted as requiring professional health care. We calculated accuracy, sensitivity, and specificity for each measure and juxtaposed the median and IQR of apps based on the 4 data sets we compared.

Comparison of Diagnostic Accuracy

We assessed the evolution of diagnostic accuracy of symptom checkers by juxtaposing median and IQRs of 3 measures of diagnostic accuracy, abbreviated as M1, M10, and M20: they are defined as the proportions of case vignettes a symptom checker assessed where it suggested the gold standard diagnosis first (M1) within the first 10 (M10) or within the first 20 (M20) diagnostic suggestions it gave. We report M1 for all 3 data sets on symptom checkers, M10 for the Hill et al [24,34] and our data set, and the M20 measure for the Semigran et al [23] data set only.

Association Between Diagnostic and Triage Accuracy

Not all but many symptom checker apps (14/22, 64% in the data set sampled by us; 11/23, 48% in the data set sampled by Semigran et al [23]; and 8/36, 22% in the data set sampled by Hill et al [24,34]) provide both diagnostic and triage advice. As users approach symptom checker apps for different reasons—for example, some people aim at self-diagnosis, whereas others are looking for guidance through the health care system [8]—we wondered whether symptom checker apps either tend to perform well or poorly in both use cases or whether apps can provide helpful information on one of these questions but not the other. Thus, we explored whether the triage accuracy of these apps was linked to their diagnostic performance by analyzing the association of triage accuracy and diagnostic accuracy with linear models, 1 for each of the 3 samples of symptom checkers, and calculated the unadjusted R^2 value as a measure of variance explained. We further determined how commonly symptom checkers erred in their triage appraisal despite suggesting the correct diagnosis first. A high proportion indicates that symptom checkers grasp the case presentation but struggle with linking the correct triage level to the case presentation, for example, by providing overcautious triage advice despite having correctly identified a diagnosis of low urgency.

Results

Study Sample

Our systematic search returned 22 symptom checkers capable of providing triage advice, 14 (64%) of which also suggested diagnoses. Approximately 23% (5/22) of symptom checkers (*K Health*, *Isabel*, *Symcat*, *Everyday Health*, and *WebMD*) differentiated only 2 triage levels—emergency care and nonemergency care—whereas the other symptom checkers included *self-care* as potential triage advice. As most of the 22 symptom checkers were unable to evaluate every case vignette (eg, as their scope was limited to adult or pediatric patients), our assessment of the apps on 45 case vignettes yielded a total of 796 unique triage evaluations (22 apps, with a median of 40 evaluations per app and an IQR of 11) and 520 unique diagnostic evaluations (14 apps, with a median of 39 evaluations per app and an IQR of 15). Tables S1-S3 in Multimedia Appendix 1 list the retrieved symptom checker apps and denote their individual triage and diagnostic performance.

Comparison of Triage Accuracy

Direct Comparison of Symptom Checkers' Triage Accuracy in 2015 and 2020

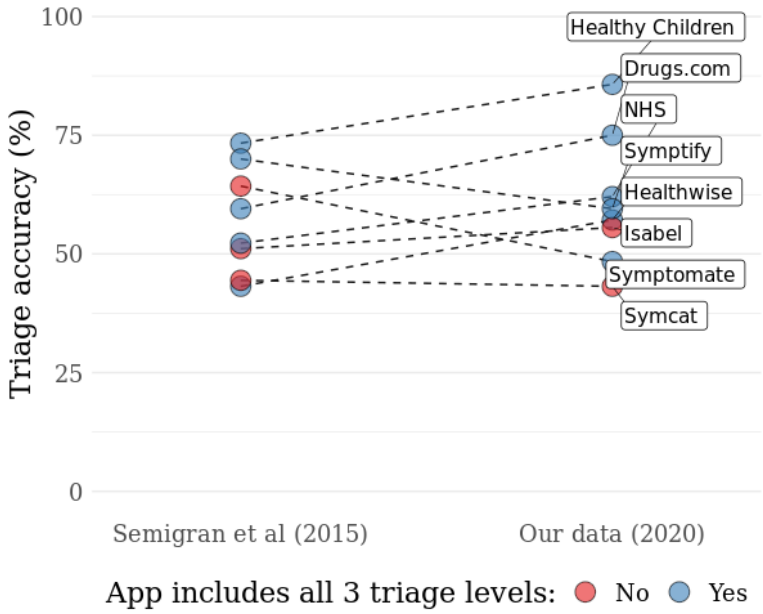
The median overall triage accuracy of all symptom checkers in our data set from 2020 (55.8%, IQR 15.1%; N=22) is close to the median triage accuracy of the apps in 2015 by Semigran et al [23] (59.1%, IQR 15.5%; N=15). The medians remain similar when excluding apps that never advise seeking self-care (Table 1). Most of those apps included both in the sample of Semigran et al [23] and our sample (5/8, 63%) improved their overall triage accuracy on the set of 45 case vignettes (Figure 1).

Table 1. Overall triage accuracy of symptom checker apps in 2015 (data from a study by Semigran et al [23]; N=15) and 2020 (data collected by us; N=22).

Sample of symptom checker apps	Overall triage accuracy			
	2015		2020	
	Values (%), median (IQR)	Values, n (%)	Values (%), median (IQR)	Values, n (%)
All triaging apps included in the respective study	59.1 (51.7-67.1)	15 (100)	55.8 (47.8-62.9)	22 (100)
Subset of apps included in both studies	55.9 (49.4-65.7)	8 (53)	58.3 (53.8-65.3)	8 (36)
Subset of apps capable of providing self-care triage advice	59.5 (53.3-70.7)	11 (73)	59.5 (50.0-64.4)	17 (77)

XSL•FO
RenderX

Figure 1. Overall triage accuracy of 8 symptom checkers included in both samples (2015 and 2020) and assessed on the same 45 case vignettes in 2015 and 2020. Data on symptom checker accuracy for 2015 are taken from a study by Semigran et al [23]. Of the 8 symptom checkers, 3 never recommended self-care as triage level (colored in red) in 2015 and 2 in 2020. One symptom checker (Symptomate) never recommended self-care in the 2015 study by Semigran et al [23] but provides such recommendations in 2020, as indicated both in our data and reported by Hill et al [24,35]. NHS: National Health Service.



Confusion Matrices for Triage Advice

The 2015 sample of 11 apps, providing triage advice and including all 3 urgency levels, more commonly erred by overtriaging a case vignette than by undertriaging (odds 2.82:1 and 110:39, respectively). The ratio of overtriage to undertriage was less for the respective sample of 17 apps in 2020 (odds 1.11:1 and 131:118, respectively). In comparison with the sample of 2015, the sample of 2020 less frequently mistook self-care cases and nonemergency cases for emergencies, whereas at the same time more often misclassified emergencies as nonemergencies (Tables 2 and 3).

Confusion matrices, including the case evaluations by those apps that did not provide triage advice on self-care, can be found in the Multimedia Appendix 1 (Tables S4 and S5). They show the same general trend as reported here.

Table 2. Confusion matrix of triage advice of 11 symptom checker apps assessed in 2015 by Semigran et al [23].

Triage recommendation provided by the symptom checker app	Gold standard solution of the triage level for the case vignette (15 case vignettes per category), n (%)		
	Emergency (n=130 evaluations)	Nonemergency (n=128 evaluations)	Self-care (n=127 evaluations)
Emergency care	103 (79.2)	41 (32)	23 (18.1)
Nonemergency	22 (16.9)	74 (57.8)	46 (36.2)
Self-care	5 (3.8)	13 (10.1)	58 (45.6)

Table 3. Confusion matrix of triage advice of 17 symptom checker apps assessed in 2020 on the same 45 case vignettes as used by Semigran et al [23] in 2015.

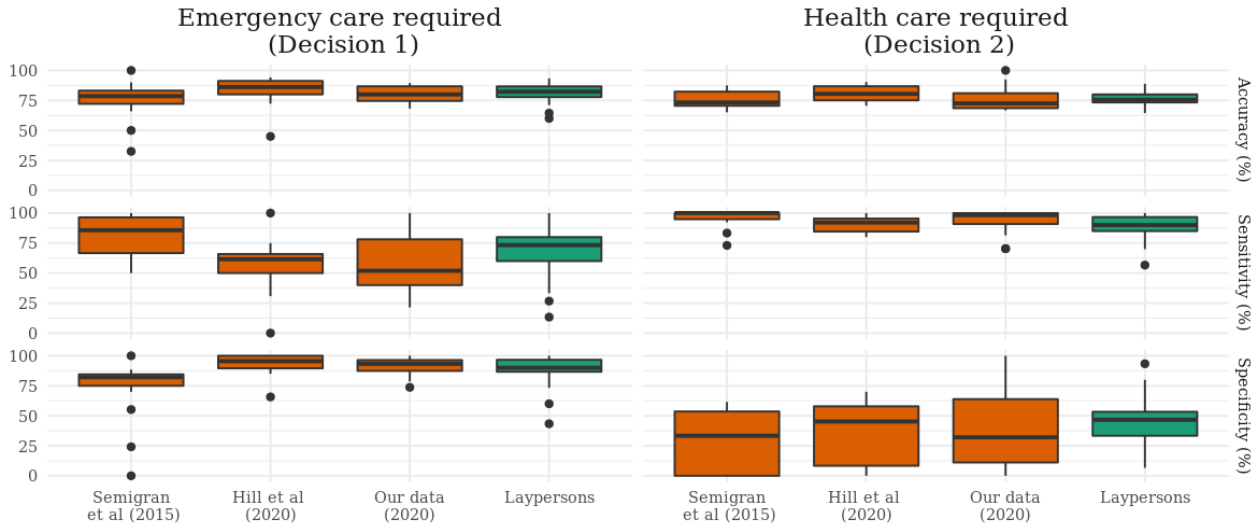
Triage recommendation provided by the symptom checker app	Gold standard solution of the triage level for the case vignette (15 case vignettes per category), n (%)		
	Emergency (n=202 evaluations)	Nonemergency (n=205 evaluations)	Self-care (n=193 evaluations)
Emergency care	116 (57.4)	26 (12.6)	6 (3.1)
Nonemergency	80 (39.6)	147 (71.7)	99 (51.2)
Self-care	6 (2.9)	32 (15.6)	88 (45.5)

XSL•FO
RenderX

Symptom Checkers' Triage Capability on Binary Triage Decisions in 2015 and 2020

The median accuracy of the apps in deciding whether emergency care is necessary (decision 1) in 2015 (78.6%, IQR 72.1%-83.1%) was similar to our re-evaluation in 2020 (80%, IQR 74.6%-86.8%; Figure 2). The same holds true regarding the median accuracy for decision 2, whether medical care should be sought or self-care is sufficient (73.3%, IQR 70.5%-82.3% vs 72.6%, IQR 68.5%-81%). Differences between the apps' triage performance in 2015 and 2020 appear when comparing the sensitivity and specificity for detecting emergencies. In 2015, the median app correctly spotted 85.7% (IQR 66.7%-96.4%) of the emergencies (sensitivity), with a median specificity of 82.1% (IQR 75%-84.5%). In comparison, in our 2020 data, the median app spotted 51.9% (IQR 40%-78.2%) of the emergencies (sensitivity) and attained a specificity of 93.3% (IQR 87.4%-96.4%). Comparing the data from studies by Semigran et al [23] and Hill et al [24,34] reveals the same trend, with a low sensitivity to identify emergencies (61.5%, IQR 50%-65.9%) and high specificity to rule them out (95.5%, IQR 89.6%-100%) in 2020. Such a trend cannot be detected regarding decision 2, whether professional medical care (health care) is required, as sensitivity and specificity scores from 2015 are close to those from 2020 (Figure 2).

Figure 2. Accuracy, sensitivity, and specificity of symptom checker apps and laypersons for 2 binary triage decisions on whether emergency care is required and whether professional medical care is required at all. Data for symptom checkers are taken from Semigran et al [23], Hill et al [24,35], and our own data collection. Data on laypersons are taken from Schmieding et al [36].



Symptom Checkers' Triage Capability on Binary Triage Decisions in 2020 Compared With Laypersons

The sample of 22 symptom checkers assessed in 2020 performed very similarly to laypersons' triage accuracy (Figure 2) [36]. However, few apps managed to outperform laypersons on binary triage decisions. Concerning decision 1, whether emergency health care should be sought or not, 18% (4/22) of apps (*Mayo*, *Ada*, *Isabel*, and *Healthwise*) showed a higher accuracy, sensitivity, and specificity than the median layperson (accuracy 82.2%, sensitivity 73.3%, and specificity 90%). Concerning decision 2, whether professional medical care should be sought, 23% (5/22) of apps (*Healthy Children*, *NHS*, *Drugs.com*, *Healthily*, and *Earlydoc*) managed to outperform the median layperson's accuracy (75.9%), sensitivity (90%), and specificity (46.7%; Table S2 in Multimedia Appendix 1).

Comparison of Diagnostic Accuracy

The 64% (14/22) of symptom checkers that provided diagnostic advice in 2020, on average, provided the correct diagnostic suggestion first (M1) for approximately half the case vignettes assessed, and two-thirds of the time, the correct diagnosis was listed among the first 10 suggestions on average (M10; Table 4). The M1 score is higher than the sample median diagnostic accuracies reported by Semigran et al [23] in 2015. In line with this, of the 7 symptom checker apps providing diagnostic suggestions and included in the samples by both Semigran et al [23] and our study, the majority (6/7) improved their M1 diagnostic accuracy (Figure 3). Hill et al [24,34] reported median diagnostic accuracy scores in 2020 closer to those of Semigran et al [23] for M1 diagnostic accuracy.

XSL•FO
RenderX

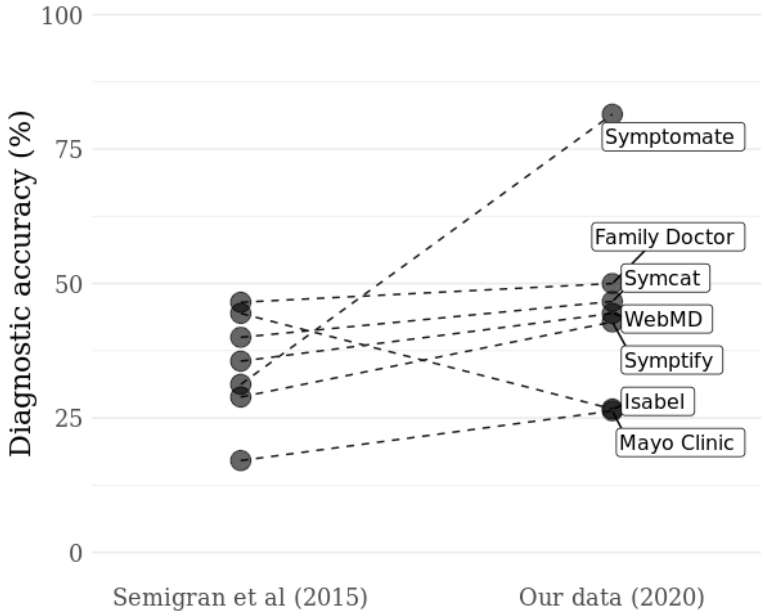
Table 4. Diagnostic accuracy of symptom checkers as reported by Semigran et al [23] in 2015, Hill et al [24,34], and our data set from 2020[a].

Metric of diagnostic accuracy	Diagnostic accuracy of symptom checkers (%), median (IQR)		
	Semigran et al [23] (n=19 apps)	Hill et al [24,34] (n=24 apps)	Our data (n=14 apps)
M1	35.5 (30.0-40.0)	34.3 (26.5-40.1)	45.5 (37.5-51.7)
M10	—[b]	59.2 (40.5-70.8)	71.1 (60.9-76.9)
M20	55.8 (45.2-73.6)	—	—

[a]Diagnostic accuracy as reported by Hill et al [24,34] is based on a different but overlapping set of case vignettes. M1, M10, and M20 abbreviate the proportion of case vignettes a symptom checker assessed where it suggested the gold standard diagnosis first (M1) within the first 10 (M10) or within the first 20 diagnostic suggestions (M20). The table displays the median and IQR values on these 3 metrics of the 3 samples of symptom checkers.

[b]Not available: Semigran et al [23] presented values only for M1, M3, and M20. Hill et al [24,34] and our data collection disregarded diagnostic suggestions beyond the first 10 suggestions.

Figure 3. Overall diagnostic accuracy (correct diagnosis listed first, M1) of 7 symptom checkers included in both samples (2015 and 2020) and assessed on the same 45 case vignettes in 2015 and 2020. Data on symptom checker accuracy for 2015 are taken from Semigran et al [23].

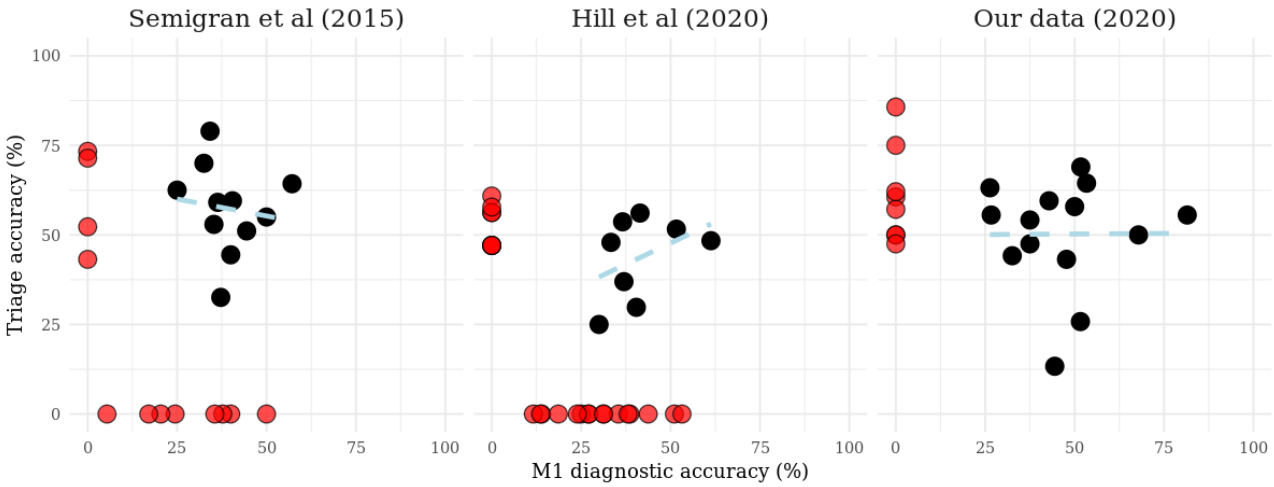


Relation Between Diagnostic and Triage Accuracy

Considering those apps that provided both diagnostic and triage advice, the proportion of wrong triage assessments when the correct diagnosis was suggested first is 37.7% (57/151) for the study by Semigran et al [23], 37.6% (88/234) in our data, and 46.4% (58/125) in the data provided by Hill et al [24,34]. Accordingly, the individual symptom checker app's top 1 diagnostic and triage accuracy does not correlate well with low unadjusted R^2 values (0.018, 0.175, and <0.001 for the Semigran et al [23], Hill et al [24,34], and our data sets, respectively; Figure 4). In the study by Semigran et al [23], most of these erroneous triage assignments were overtriage errors (52/57, 91%), whereas this proportion was lower in our data (46/88, 52%) and in that of the Hill et al [24,34] study (34/58, 58%). Concerning those evaluations where symptom checkers got the diagnosis right but allocated the wrong triage levels, many of the errors were because of a false appraisal of whether emergency care was necessary or not (Semigran et al [23] study: 29/57, 51%; Hill et al [24,34] study: 18/58, 31%; and our data: 24/88, 27%).

XSL•FO
RenderX

Figure 4. Association between M1 diagnostic accuracy (proportion of case vignettes to which the app provided the correct diagnosis first, as percentage) and triage accuracy. Every dot represents a symptom checker app. Red dots represent apps that provide either only triage or only diagnostic advice. Data for symptom checkers are taken from studies by Semigran et al [23], Hill et al [24,35], and our own data collection.



Discussion

Principal Findings

Evolution of Triage Capability of Symptom Checker Apps

Our study assesses how the triage and diagnostic capability of symptom checkers evolved from 2015 to 2020. A direct comparison between the data from Semigran et al [23] and the data collected by us in 2020 indicates that overall triage accuracy has not improved over the course of 5 years with respect to the same set of case vignettes. This holds true even when we look beyond the overall triage accuracy rate with 3 triage levels (emergency, nonemergency, and self-care) and instead assess the apps' accuracy to advise on seeking emergency versus no emergency care (decision 1) or seeking medical care versus not care at all (decision 2).

However, the pattern where symptom checker apps perform well and poorly seems to have changed. In 2015, apps were more risk averse (ie, they detected emergencies reliably and tended to overtriage less urgent cases), whereas both our data and the data from Hill et al [24,34] show that in 2020, apps tended to be less risk averse and missed more emergencies. The ability to reliably detect emergencies (the sensitivity to identify emergencies) can be considered the most important metric for assessing a symptom checker's safety. On the other hand, the ability to correctly spot those cases where self-care is sufficient (ie, the specificity to correctly rule out cases when professional health care is not necessary) can be considered the most important metric for assessing the usefulness of an app to both its users and the health care system, as this is the most difficult decision for laypersons [36], providing a great potential to disburden health care services. According to the data of both Hill et al [24,34] and our own study, symptom checkers still did not perform particularly well on both metrics in 2020. Comparing the distribution of triage errors (Table 3) with that of medical laypersons, as reported by Schmieding et al [36], we observe that the triage behavior of symptom checkers and medical laypersons have aligned. Thus, we consider it a pressing question whether symptom checkers can enhance laypersons' decision-making when both their accuracy and direction of errors are similar. The importance of this question is supported by an experimental study demonstrating that most laypersons do not change their triage assessment after an internet search, and when they do, this change is as commonly correct as it is incorrect [41].

However, it must be noted that among our sample of apps, some defy the general trend and show high rates of accuracy, sensitivity, and specificity for either but not both of the binary triage decisions, indicating that they are potentially beneficial to their users when approached with the appropriate question (Table S2, Multimedia Appendix 1).

Evolution of Diagnostic Capability of Symptom Checker Apps

The rate of correct diagnoses being presented first (M1 diagnostic accuracy) by symptom checkers is still low, with a median of <50%. Only 2 symptom checkers (*Everyday Health* and *Symptomate*) in our sample achieved an M1 diagnostic value near the M1 diagnostic accuracy of physicians on these case vignettes, reported at 72.1% by Semigran et al [42] in a study from 2016 (Table S3, Multimedia Appendix 1). However, it must be noted that both apps did not evaluate all 45 case vignettes, and thus, their observed diagnostic accuracy might be skewed by selection bias. However, in contrast to the evolution of triage capability, data from a study by Hill et al [24,34] and our own data hint at a slight improvement of diagnostic accuracy, as the correct diagnosis is now more commonly included in the top 10 suggestions (M10) than in the first 20 suggestions in 2015. However, it must be considered that these case vignettes have been publicly available, and therefore, app developers potentially trained their apps' algorithms on these cases. Hence, the diagnostic accuracy for previously unpublished case vignettes might have increased differently or not all. In addition, as users are presumably most

XSL•FO
RenderX

affected by the first or first few diagnostic suggestions, we question whether the observed increase in diagnostic capability already translates into an additional benefit noticeable to users.

Association Between Symptom Checker Apps' Diagnostic and Triage Capability

All 3 data sets assessing symptom checker performance considered in this study indicate that correct diagnostic evaluation does not reliably translate into correct triage evaluation; that is, apps capable of diagnosing correctly are not necessarily good at triaging. Given that a large proportion of triage advice was incorrect even when the app correctly diagnosed the case (between 37% and 46%), a cause for the divergence of triage and diagnostic might be that app developers assign a wrong triage level to their diagnoses. However, as symptom checkers are developed by companies and agencies from different countries with potentially very different health care systems, the assignment of a specific triage level to a diagnosis might be wrong in one health care setting but correct in another. We observed that a considerable proportion of correctly diagnosed but incorrectly triaged case vignettes were incorrectly classified as (not) constituting an emergency. As we deem the variation between health care systems concerning the definition of what constitutes an emergency low, triage advice tailored to a specific health care system cannot be the only explanation for the observed discrepancies between triage and diagnostic capability.

Limitations

The assessment of symptom checker performance in our study is based on clinical vignettes and thus comes with important limitations, which our study shares with other case vignette-based approaches to assess symptom checkers: patient users might enter their complaints less reliably and more ambiguously into apps compared with the mock users who enter information from clinical vignettes in a more structured fashion. A study by Jungmann et al [43] shows low interrater reliability among laypersons in entering information into symptom checker apps. Thus, vignette-based studies potentially have a poor ecological validity (ie, transferability to the real-world setting) and might overestimate the accuracy of symptom checkers when used by their target users—laypersons—in a real-world setting as they do not account for users' variable capability to enter their symptoms without making errors.

Furthermore, researchers assign a correct solution (gold standard diagnosis and triage level) to each case vignette. However, there might be >1. These case vignettes commonly represent the first presentation of new complaints of a fictitious patient, and thus—similar to the real clinical setting—to definitely determine the correct diagnosis and most appropriate triage level, additional information might be required, such as imaging or laboratory findings. Thus, at the time of initial presentation, which is also the time when symptom checkers tend to be used, multiple options might be considered correct when more predictive information is not yet available.

In addition, even when symptom checkers achieve high accuracy, their true value to the users can only be fully estimated when taking into account the users' own appraisal, prior knowledge, and trust in the symptom checker [41,44,45]. Thus, an evaluation of symptom checkers with case vignettes alone is a useful but only a first step to identify the best symptom checkers; in a second step, the best-in-class apps should then be further evaluated with study designs where patients enter their own complaints [46-48], and patient-reported outcomes and experience measures should be brought into focus.

Despite the limitations of vignette-based audit studies, we are convinced they will remain essential, as they provide a means for quick and potentially automatable evaluations of symptom checkers. With symptom checker software being updated on a regular basis and new apps continuously becoming available, we consider the vignette-based approach a necessary complement to more informative but also more costly and lengthy clinical studies. Thus, we advocate for the further development of the key resources of such studies—the clinical vignettes. We suggest the creation of a repository of vignettes not only based on real patient histories but also refined by a test–theoretical perspective and annotated by machine-readable codes (such as Systematized Nomenclature of Medicine–Clinical Terms semantic tags) to pave the way for automating the evaluation of those symptom checkers providing a suitable application programming interface.

Apart from the limitations deriving from the use of vignettes, our study comes with an additional set of specific limitations. As symptom checkers appear and disappear, the sample of retrievable symptom checkers varies over time. Of the 15 triaging symptom checkers assessed in 2015, only 8 (53%) were retrievable in 2020. In addition, although an implicit consensus on defining symptom checkers by their function emerges, it is difficult to definitely determine whether a *tool* can be considered a general-purpose symptom checker app and thus be included in studies. For example, the tool *Healthy Children* was considered a symptom checker app by Semigran et al [23] and Hill et al [24], although it merely presents a list of advice and descriptions of common diagnoses associated with single chief complaints without performing an input-based assessment of a patient's complaints as most other tools do. For the purpose of comparability, we included *Healthy Children* and similar tools in our study as well, although we are aware that other researchers disregard these tools in their studies [32], arguably as such tools have not much in common with smartphone apps or web-based applications more reactive to user input and built on computational rather a tree-based algorithms, except their shared use case. Consequently, any attempt to assess the presumed population of symptom checker apps faces the difficulty that what constitutes a symptom checker is ill defined. To avoid rarely used, potentially poorly performing apps distorting the results, we excluded apps with few downloads or below a certain threshold of user rating scores. By doing so, the inclusion criteria for our study were more strict than in the study by Semigran et al [23]; however, the limitations certainly remain.

The heterogeneous definitions of triage levels potentially pose an additional, important limitation for all symptom checker comparison studies. Some studies on symptom checkers use only 2 triage levels [25], whereas others use as many as 6 [33]. By defining 2 binary metrics of triage accuracy rather than just

XSL•FO
RenderX

1, we could mitigate this limitation but only partially, as triage-level definitions can also be incongruent between studies as, for instance, the definition of *urgent care* by Hill et al [24] mostly covers what Semigran et al [23] define as *nonemergency* care but partially overlaps with the definition of *emergency care* by Semigran et al [23] as well. Thus, we recommend that future studies that aim to compare the triage capability of symptom checkers competitively also include binary triage measures that resemble real-life decisions (eg, should I call an ambulance or not and is self-care sufficient and safe?), in addition to more compartmentalized classifications of triage levels that can be tailored for the local health care system at best and arbitrary at worst. Although more compartmentalized classifications make a meaningful comparison between symptom checkers with different triage-level definitions more difficult, this approach does acknowledge that the potential real-world benefit of symptom checkers also lies in their ability to guide their users through the health care system with advice that is as specific as possible.

Finally, we consider our greatest limitation that a single mock user, a nonnative English speaker, compiled our data for symptom checker performance in 2020. We tried to minimize but certainly did not eliminate the risk that our comparative analyses are influenced by this by (1) complementing the case vignettes in advance with additional information we anticipated symptom checkers would prompt the mock user to provide and (2) defining rules on how to handle ambiguities in symptom checkers' questions or in the vignettes.

Conclusions

Assessing the capabilities of symptom checkers in a transparent and reproducible manner is challenging but necessary to gather independent and non–industry-funded evidence on these increasingly popular decision support tools for patients and laypersons. Our study compares 3 data sets on symptom checkers' diagnostic and triage performance, 1 with data from 2015 and 2 with data from 2020. Taken together, they suggest that symptom checkers' triage performance has, on average, not improved over the course of 5 years, and it potentially even decreased in the most important use cases (safe advice on when emergency care is required and when no health care is needed for the moment). Few highly performing apps managed to provide more reliable triage advice than an average layperson in one of those important 2 use cases. However, no symptom checker outperformed the laypersons in both use cases, and in general, symptom checkers' triage behavior has become more similar to that of laypersons. Although some apps are good at both triaging and diagnosis, no general association between an app's triage and diagnostic ability exists to date. In addition, the accuracy of advice does not only vary considerably between symptom checkers but also within a given symptom checker, as it may prove more reliable in appraising some categories of diseases than others [48]. Taken together, these findings highlight that the current value of symptom checkers heavily depends not only on the app system but also on the question (use case) with which it is approached, for instance, whether to seek care or, if so, where or for what. Thus, medical laypersons seeking useful decision support from symptom checkers face the complexity of choosing which tool to use for what. To aid the public in taking advantage of this emerging technology, future research should develop resources (eg, repositories of case vignettes) and frameworks with which symptom checkers' performance can be evaluated continuously and independently. Together with research findings on how users integrate symptom checker's advice into their decision-making, findings on these decision aid's capabilities can provide valuable guidance as to when and to whom their use can be recommended.

Acknowledgments

No funding was required to collect and analyze the data. The authors acknowledge support from the German Research Foundation (Deutsche Forschungsgemeinschaft) and the Open Access Publication Fund of Charité Universitätsmedizin Berlin.

Authors' Contributions

MLS conceived the study, adapted the case vignettes, conducted the analyses, created most figures and all tables, and wrote the first draft of the paper. KS assisted with the case vignette adaptations and manuscript development. MK adapted and created some figures. MK, SSN, FB, and MAF provided critical input and advised on the study design, the analysis methods, and the drafts of the paper. All authors accept full responsibility for the final version of the paper. The study builds on data collected for the master's thesis by Anna Dames as a partial requirement for her MSc degree in Human Factors in the Department of Psychology and Ergonomics (Institut für Psychologie und Arbeitswissenschaft) at Technische Universität Berlin.

Conflicts of Interest

All authors have completed the International Committee of Medical Journal Editors uniform disclosure form and declare no support from any organization for the submitted work, no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years, and no other relationships or activities that could appear to have influenced the submitted work. MLS reports that he was an employee of medx GmbH (now Ada Health GmbH) in 2014 and 2015. The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported, that no important aspects of the study have been omitted and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

XSL•FO
RenderX

Multimedia Appendix 1
Supplementary tables and details on methods.
[DOCX File , 97 KB - jmir_v24i5e31810_app1.docx]

References

1. Lunn PD, Timmons S, Julienne H, Belton CA, Barjaková M, Lavin C, et al. Using decision aids to support self-isolation during the COVID-19 pandemic. Psychol Health 2021 Feb;36(2):195-213. [doi: 10.1080/08870446.2020.1849701] [Medline: 33210950]
2. Zhang MW, Chow A, Ho RC, Smith HE. An overview of commercially available apps in the initial months of the COVID-19 pandemic. Front Psychiatry 2021 Apr 15;12:557299 [FREE Full text] [doi: 10.3389/fpsyt.2021.557299] [Medline: 33935816]
3. Judson TJ, Odisho AY, Neinstein AB, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
4. Alanzi T. A review of mobile applications available in the App and Google play stores used during the COVID-19 outbreak. J Multidiscip Healthc 2021 Jan 12;14:45-57 [FREE Full text] [doi: 10.2147/JMDH.S285014] [Medline: 33469298]
5. World Health Organization. Applications of artificial intelligence for health. In: Ethics and governance of artificial intelligence for health: WHO guidance. Geneva, Switzerland: World Health Organization; Jun 28, 2021:15.
6. Farmer SE, Bernardotto M, Singh V. How good is Internet self-diagnosis of ENT symptoms using Boots WebMD symptom checker? Clin Otolaryngol 2011 Oct;36(5):517-518. [doi: 10.1111/j.1749-4486.2011.02375.x] [Medline: 22032458]
7. EPatient survey 2020. Health & Care Management. 2020. URL: https://www.hcm-magazin.de/epatient-survey-2020-digital-health-studie/150/10992/407743 [accessed 2021-03-06]
8. Meyer AN, Giardina TD, Spitzmueller C, Shahid U, Scott TM, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: cross-sectional survey study. J Med Internet Res 2020 Jan 30;22(1):e14679 [FREE Full text] [doi: 10.2196/14679] [Medline: 32012052]
9. Elliot AJ, Kara EO, Loveridge P, Bawa Z, Morbey RA, Moth M, et al. Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system. Epidemiol Infect 2015 Dec;143(16):3416-3422. [doi: 10.1017/S0950268815000503] [Medline: 25858297]
10. Mehl A, Bergey F, Cawley C, Gilsdorf A. Syndromic surveillance insights from a symptom assessment app before and during COVID-19 measures in Germany and the United Kingdom: results from repeated cross-sectional analyses. JMIR Mhealth Uhealth 2020 Oct 09;8(10):e21364 [FREE Full text] [doi: 10.2196/21364] [Medline: 32997640]
11. Medical symptom checker. Sutter Health. URL: https://www.sutterhealth.org/health/symptom-checker [accessed 2021-11-13]
12. Sutter health teams up with ADA health to improve patient care by delivering on-demand healthcare guidance. Sutter Health. 2019 Feb 11. URL: https://vitals.sutterhealth.org/sutter-health-teams-up-with-ada-health-to-improve-patient-care-by-delivering-on-demand-healthcare-guidance/ [accessed 2021-11-13]
13. Check your symptoms: who has the problem? Kaiser Permanente. URL: https://healthy.kaiserpermanente.org/health-wellness/health-encyclopedia/he.check-your-symptoms.hwsxchk [accessed 2021-11-13]
14. Laukka E, Kujala S, Gluschkoff K, Kanste O, Hörhammer I, Heponiemi T. Leaders' support for using online symptom checkers in Finnish primary care: survey study. Health Informatics J 2021;27(4):14604582211052259 [FREE Full text] [doi: 10.1177/14604582211052259] [Medline: 34821152]
15. Aboueid S, Meyer S, Wallace JR, Mahajan S, Chaurasia A. Young adults' perspectives on the use of symptom checkers for self-triage and self-diagnosis: qualitative study. JMIR Public Health Surveill 2021 Jan 06;7(1):e22637 [FREE Full text] [doi: 10.2196/22637] [Medline: 33404515]
16. Kujala S, Hörhammer I, Hänninen-Ervasti R, Heponiemi T. Health professionals' experiences of the benefits and challenges of online symptom checkers. Stud Health Technol Inform 2020 Jun 16;270:966-970. [doi: 10.3233/SHTI200305] [Medline: 32570525]
17. Using technology to ease the burden on primary care. Healthwatch Enfield. 2019 Jan 19. URL: https://www.healthwatch.co.uk/sites/healthwatch.co.uk/files/reports-library/20190122_Enfield_%20Using%20technology%20to%20ease%20the%20burden%20on%20primary%20care.pdf [accessed 2021-03-13]
18. Fraser H, Coiera E, Wong D. Safety of patient-facing digital symptom checkers. Lancet 2018 Nov 24;392(10161):2263-2264. [doi: 10.1016/S0140-6736(18)32819-8] [Medline: 30413281]
19. Berner ES, Webster GD, Shugerman AA, Jackson JR, Algina J, Baker AL, et al. Performance of four computer-based diagnostic systems. N Engl J Med 1994 Jun 23;330(25):1792-1796. [doi: 10.1056/NEJM199406233302506] [Medline: 8190157]
20. Bond WF, Schwartz LM, Weaver KR, Levick D, Giuliano M, Graber ML. Differential diagnosis generators: an evaluation of currently available computer programs. J Gen Intern Med 2012 Feb;27(2):213-219 [FREE Full text] [doi: 10.1007/s11606-011-1804-8] [Medline: 21789717]
21. Yager J, Linn LS, Leake B, Gastaldo G, Palkowski C. Initial clinical judgments by internists, family physicians, and psychiatrists in response to patient vignettes: I. Assessment of problems and diagnostic possibilities. Gen Hosp Psychiatry 1986 May;8(3):145-151. [doi: 10.1016/0163-8343(86)90072-1] [Medline: 3710146]

XSL•FO
RenderX

22. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. JAMA 2000 Apr 05;283(13):1715-1722. [doi: 10.1001/jama.283.13.1715] [Medline: 10755498]
23. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015 Jul 08;351:h3480 [FREE Full text] [doi: 10.1136/bmj.h3480] [Medline: 26157077]
24. Hill MG, Sim M, Mills B. The quality of diagnosis and triage advice provided by free online symptom checkers and apps in Australia. Med J Aust 2020 Jun;212(11):514-519. [doi: 10.5694/mja2.50600] [Medline: 32391611]
25. Ćirković A. Evaluation of four artificial intelligence-assisted self-diagnosis apps on three diagnoses: two-year follow-up study. J Med Internet Res 2020 Dec 04;22(12):e18097 [FREE Full text] [doi: 10.2196/18097] [Medline: 33275113]
26. Berry AC, Cash BD, Mulekar MS, Wang B, Melvin A, Berry BB. Symptom checkers vs. doctors, the ultimate test: a prospective study of patients presenting with abdominal pain. Gastroenterology 2017 Apr;152(5):S852-S853. [doi: 10.1016/s0016-5085(17)32937-2]
27. Yu SW, Ma A, Tsang VH, Chung LS, Leung SC, Leung LP. Triage accuracy of online symptom checkers for Accident and Emergency Department patients. Hong Kong J Emerg Med 2019 Apr 16;27(4):217-222. [doi: 10.1177/1024907919842486]
28. Gottliebsen K, Petersson G. Limited evidence of benefits of patient operated intelligent primary care triage tools: findings of a literature review. BMJ Health Care Inform 2020 May;27(1):e100114 [FREE Full text] [doi: 10.1136/bmjhci-2019-100114] [Medline: 32385041]
29. Wallace W, Chan C, Chidambaram S, Hanna L, Iqbal FM, Acharya A, et al. The diagnostic and triage accuracy of digital and online symptom checker tools: a systematic review. medRxiv (forthcoming) 2021. [doi: 10.1101/2021.12.21.21268167]
30. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. J Med Internet Res 2017 Jun 13;19(6):e202 [FREE Full text] [doi: 10.2196/jmir.6755] [Medline: 28611017]
31. Chambers D, Cantrell AJ, Johnson M, Preston L, Baxter SK, Booth A, et al. Digital and online symptom checkers and health assessment/triage services for urgent health problems: systematic review. BMJ Open 2019 Aug 01;9(8):e027743 [FREE Full text] [doi: 10.1136/bmjopen-2018-027743] [Medline: 31375610]
32. Ceney A, Tolond S, Glowinski A, Marks B, Swift S, Palser T. Accuracy of online symptom checkers and the potential impact on service utilisation. PLoS One 2021 Jul 15;16(7):e0254088 [FREE Full text] [doi: 10.1371/journal.pone.0254088] [Medline: 34265845]
33. Gilbert S, Mehl A, Baluch A, Cawley C, Challiner J, Fraser H, et al. How accurate are digital symptom assessment apps for suggesting conditions and urgency advice? A clinical vignettes comparison to GPs. BMJ Open 2020 Dec 16;10(12):e040269 [FREE Full text] [doi: 10.1136/bmjopen-2020-040269] [Medline: 33328258]
34. Hill MG. Appraisal of free online symptom checkers and applications for self-diagnosis and triage: an Australian evaluation. Edith Cowan University. 2020. URL: https://ro.ecu.edu.au/theses/2311/ [accessed 2021-07-03]
35. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Data set supplementing "Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study". Zenodo. 2021 Jan 21. URL: https://zenodo.org/record/4454538/export/hx [accessed 2021-01-21]
36. Schmieding ML, Mörgeli R, Schmieding MA, Feufel MA, Balzer F. Benchmarking triage capability of symptom checkers against that of medical laypersons: survey study. J Med Internet Res 2021 Mar 10;23(3):e24475 [FREE Full text] [doi: 10.2196/24475] [Medline: 33688845]
37. Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Blazer F, Feufel M. Data set on accuracy of symptom checker apps in 2020. Zenodo. 2022 Feb 12. URL: https://zenodo.org/record/6054093 [accessed 2022-02-12]
38. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2021. URL: https://www.R-project.org/ [accessed 2021-05-14]
39. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov 21;4(43):1686. [doi: 10.21105/joss.01686]
40. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Cham, Switzerland: Springer; 2016.
41. Levine DM, Mehrotra A. Assessment of diagnosis and triage in validated case vignettes among nonphysicians before and after internet search. JAMA Netw Open 2021 Mar 01;4(3):e213287 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.3287] [Medline: 33779741]
42. Semigran HL, Levine DM, Nundy S, Mehrotra A. Comparison of physician and computer diagnostic accuracy. JAMA Intern Med 2016 Dec 01;176(12):1860-1861. [doi: 10.1001/jamainternmed.2016.6001] [Medline: 27723877]
43. Jungmann SM, Klan T, Kuhn S, Jungmann F. Accuracy of a chatbot (Ada) in the diagnosis of mental disorders: comparative case study with lay and expert users. JMIR Form Res 2019 Oct 29;3(4):e13863 [FREE Full text] [doi: 10.2196/13863] [Medline: 31663858]
44. Woodcock C, Mittelstadt B, Busbridge D, Blank G. The impact of explanations on layperson trust in artificial intelligence-driven symptom checker apps: experimental study. J Med Internet Res 2021 Nov 03;23(11):e29386 [FREE Full text] [doi: 10.2196/29386] [Medline: 34730544]

XSL•FO
RenderX

45. Winn AN, Somai M, Fergestrom N, Crotty BH. Association of use of online symptom checkers with patients' plans for seeking care. JAMA Netw Open 2019 Dec 02;2(12):e1918561 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18561] [Medline: 31880791]
46. Martin SS, Quaye E, Schultz S, Fashanu OE, Wang J, Saheed MO, et al. A randomized controlled trial of online symptom searching to inform patient generated differential diagnoses. NPJ Digit Med 2019 Nov 11;2:110 [FREE Full text] [doi: 10.1038/s41746-019-0183-0] [Medline: 31728417]
47. Knitza J, Mohn J, Bergmann C, Kampylafka E, Hagen M, Bohr D, et al. Accuracy, patient-perceived usability, and acceptance of two symptom checkers (Ada and Rheport) in rheumatology: interim results from a randomized controlled crossover trial. Arthritis Res Ther 2021 Apr 13;23(1):112 [FREE Full text] [doi: 10.1186/s13075-021-02498-8] [Medline: 33849654]
48. Hennemann S, Kuhn S, Witthöft M, Jungmann SM. Diagnostic performance of an app-based symptom checker in mental disorders: comparative study in psychotherapy outpatients. JMIR Ment Health 2022 Jan 31;9(1):e32832 [FREE Full text] [doi: 10.2196/32832] [Medline: 35099395]

Edited by A Mavragani; submitted 07.07.21; peer-reviewed by J Knitza, M Hill, C Lowe; comments to author 07.08.21; revised version received 19.11.21; accepted 30.01.22; published 10.05.22.

Please cite as:
Schmieding ML, Kopka M, Schmidt K, Schulz-Niethammer S, Balzer F, Feufel MA
Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation
J Med Internet Res 2022;24(5):e31810
URL: https://www.jmir.org/2022/5/e31810
doi:10.2196/31810
PMID:35536633

©Malte L Schmieding, Marvin Kopka, Konrad Schmidt, Sven Schulz-Niethammer, Felix Balzer, Markus A Feufel. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 10.05.2022. 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.

XSL•FO
RenderX

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX